



TESIS DOCTORAL

MODALIDAD DE COMPENDIO DE  
ARTÍCULOS

EVALUACIÓN FORENSE DE LA  
CREDIBILIDAD DEL TESTIMONIO Y  
SINTOMATOLOGÍA INTERNALIZANTE  
EN DELITOS COMETIDOS EN  
LA ESFERA PRIVADA

Bárbara González Amado

PROGRAMA DE DOUTORAMENTO EN PSICOLOXÍA DO TRABALLO  
E AS ORGANIZACIÓNS, XURÍDICA-FORENSE E DO CONSUMIDOR E  
USUARIO (RD 99/2011)

FACULTADE DE PSICOLOXÍA

SANTIAGO DE COMPOSTELA

2017





TESIS DOCTORAL

MODALIDAD DE COMPENDIO DE  
ARTÍCULOS

EVALUACIÓN FORENSE DE LA  
CREDIBILIDAD DEL TESTIMONIO Y  
SINTOMATOLOGÍA INTERNALIZANTE  
EN DELITOS COMETIDOS EN  
LA ESFERA PRIVADA

Fdo.: .....

Bárbara González Amado

PROGRAMA DE DOUTORAMENTO EN PSICOLOXÍA DO TRABALLO  
E AS ORGANIZACIÓNS, XURÍDICA-FORENSE E DO CONSUMIDOR E  
USUARIO (RD 99/2011)

FACULTADE DE PSICOLOXÍA

SANTIAGO DE COMPOSTELA

2017





**AUTORIZACIÓN DO DIRECTOR/TITOR E**

**DIRECTORA DA TESE:**

**AVALIACIÓN FORENSE DA CREDIBILIDADE DAS TESTEMUÑAS E  
SINTOMATOLOXÍA INTERNALIZANTE EN DELITOS COMETIDOS NA ESFERA  
PRIVADA**

D. Ramón Arce Fernández, Catedrático de Psicoloxía Xurídica e Forense da Universidade de Santiago de Compostela e Dna. Francisca Fariña Rivera, Catedrática de Psicoloxía Básica e Psicoloxía Xurídica do Menor da Universidade de Vigo,

**INFORMAN**

*Que a presente tese, correspóndese co traballo realizado por Dna. Bárbara González Amado, baixo a nosa dirección, e autorizamos á presentación da tese indicada, considerando que reúne os requisitos esixidos no artigo 33 do regulamento de Estudos de Doutoramento da USC, e que como director e directora desta non incurrimos nas causas de abstención establecidas na lei 40/2015.*

Asinado. Ramón Arce Fernández

Asinado. Francisca Fariña Rivera



Prof. Dr. D. Ramón Arce Fernández e a Profa. Dra. Dna. Francisca Fariña Rivera, como director/titor e directora da tese titulada: AVALIACIÓN FORENSE DA CREDIBILIDADE DAS TESTEMUÑAS E SINTOMATOLOXÍA INTERNALIZANTE EN DELITOS COMETIDOS NA ESFERA PRIVADA.

Pola presente **DECLARAMOS:**

Que a tese presentada por Dona Bárbara González Amado é idónea para ser presentada, de acordo co artigo 41 do *Regulamento de Estudos de Doutoramento*, pola modalidade de compendio de ARTIGOS, nos que o doutorando tivo participación no peso da investigación e a súa contribución foi decisiva para levar a cabo este traballo.

E que está en coñecemento dos coautores, tanto doutores como non doutores, participantes nos artigos, que ningún dos traballos reunidos nesta tese serán presentados por ningún deles noutra tese de Doutoramento, o que asino baixo a miña responsabilidade.

Santiago de Compostela, a 9 de Maio de 2017

Asinado. Ramón Arce Fernández

Asinado. Francisca Fariña Rivera





## AGRADECEMENTOS

Agradecer, en primeiro lugar, ó meu tutor e director, Ramón Arce, e directora de tese, Francisca Fariña, pola confianza depositada en min durante estes anos e por todas as horas adicadas a transmitirme os seus coñecementos. A Mercedes Novo e Loli Seijo, polos seus consellos e pola súa inesgotable capacidade para facerme crer en min mesma. A todos eles darlles as gracias porque aquí comeza unha traxectoria profesional, cunha mochila que porto ás miñas costas cargada de aprendizaxes, e que non sería posible sen a súa labor.

Non quero esquecerme do profesor Jesús Salgado, a quen teño que agradecer a súa inestimable axuda durante os momentos de dúbida e, por suposto, por ser quén de contaxiarme a complexa beleza do mundo do meta-análise. Agradezo ó profesor Rui Abrunhosa a súa hospitalidade, profesionalidade e a atención que me brindou sempre que lle foi requerida.

As miñas compañeiras e compañeiros da Unidade de Psicoloxía Forense, agradecerlles as xornadas de choiva de ideas e de reflexión. Parte da mochila que levo tamén é labor vosa.

Ó meu pai e á miña nai, ós que lles debo todo o que son, agradezo o seu apoio incondicional e as súas verbas sempre oportunas que me mantiveron en pé. Ó seu espírito de loita é o que me empurra a seguir cara adiante. A mochila está cargada de todas as vosas ensinanzas.

A ti, meu compañeiro de vida, por servirme de reflexo e de guía.





# EVALUACIÓN FORENSE DE LA CREDIBILIDAD DEL TESTIMONIO Y SINTOMATOLOGÍA INTERNALIZANTE EN DELITOS COMETIDOS EN LA ESFERA PRIVADA

## RESUMEN

La evaluación psicológica del testimonio de la víctima constituye un elemento de prueba crucial en los delitos cometidos en la esfera privada pues, con frecuencia, es la única evidencia disponible. La técnica más utilizada para evaluar la realidad del testimonio es el Análisis de Contenido Basado en Criterios (CBCA), que se sustenta en la *Hipótesis Undeutsch*: las memorias de eventos realmente experimentados difieren en contenido y calidad (criterios de realidad) de las memorias de hechos fabricados. A pesar de su extendido uso en la justicia, en el ámbito científico goza de opiniones y resultados aparentemente contradictorios respecto de su eficacia en la discriminación de ambas memorias y, por tanto, de su aplicabilidad al contexto forense. Por su parte, su análisis, conjuntamente al de las secuelas internalizantes consecuencia de la victimización de abuso sexual infantil (ASI) o adolescente (ASA), corroboran la veracidad de la declaración y contribuyen a la suficiencia de la prueba. Ante este estado de la cuestión, nos planteamos poner a prueba la Hipótesis con los criterios de realidad. Asimismo, se evaluará la admisibilidad de la técnica en la justicia a partir del estudio de los *Criterios Daubert* y jurisprudenciales. Además, nos planteamos cuantificar la huella (trastorno depresivo y/o ansioso) consecuencia de la victimización y estimar su probabilidad. Los resultados mostraron que la Hipótesis Undeutsch y el CBCA son válidos, discriminan entre memorias vividas y fabricadas, en diferentes poblaciones (i.e., menores y adultos) y contextos. Aunque la técnica cumple los criterios Daubert, la tasa de error de clasificación de declaraciones falsas como reales en estudios de campo no es admisible (9.8% vs. 28.5%, en menores y adultos, respectivamente). Es más, se desconoce el criterio de decisión estricto para la correcta clasificación, por tanto, no es directamente extrapolable a la práctica forense. Asimismo, el sistema categorial de análisis de contenido no es metódico, en consecuencia, no válido. Por otro lado, víctimas de ASI/ASA tienen una probabilidad de .70 de desarrollar un trastorno internalizante y un 30% de daño en secuela general, depresión y ansiedad. Concretamente, la probabilidad de un daño crónico (distimia) fue significativamente superior a un daño más severo. En conclusión, la victimización de ASI/ASA repercute en el ajuste psicológico en la adultez, debiendo cuantificarse a partir de la evaluación actividad global de la víctima (EEAG).

**Palabras clave:** Hipótesis Undeutsch; criterios de realidad; CBCA; huella psicológica; abuso sexual infantil y adolescente.



# AVALIACIÓN FORENSE DA CREDIBILIDADE DAS TESTEMUÑAS E SINTOMATOLOXÍA INTERNALIZANTE EN DELITOS COMETIDOS NA ESFERA PRIVADA

## RESUMO

A avaliación psicolóxica da testemuña da vítima constitúe un elemento de proba crucial nos delitos cometidos na esfera privada pois, con frecuencia, é a única evidencia dispoñible. A técnica máis utilizada para avaliar a realidade da testemuña é o Análise de Contido Baseado en Criterios (CBCA), que se sustenta na *Hipótese Undeutsch*: as memorias de feitos realmente experimentados difiren en contido e calidade (criterios de realidade) das memorias de feitos fabricados. A pesares do seu estendido uso na xustiza, na esfera científica goza de opinións e resultados aparentemente contraditorios respecto da súa eficacia na discriminación de ámbalas dúas memorias e, polo tanto, da súa aplicabilidade ó contexto forense. Pola súa banda, o seu análise conxuntamente co das secuelas internalizantes consecuencia da vitimización de abuso sexual infantil (ASI) e adolescente (ASA), corroboran a veracidade da declaración e contribúen á suficiencia da proba. Ante este estado da cuestión, plantexámonos poñer a proba a Hipótese cos criterios de realidade. Así mesmo, avaliarase a admisibilidade da técnica na xustiza a partir do estudio dos *Criterios Daubert* e xurisprudenciais. Ademais, plantexámonos cuantificar a pegada (trastorno depresivo e/ou ansioso) consecuencia da vitimización y estimar a súa probabilidade. Os resultados amosaron que a Hipótese Undeutsch e o CBCA son válidos, discriminan entre memorias vividas e fabricadas, en diferentes poboacións (i.e., menores e adultos) e contextos. Aínda que a técnica cumpre os criterios Daubert, a taxa de error de clasificación de declaracións falsas coma reais en estudos de campo non é admisible (9.8% vs. 28.5% en menores e adultos, respectivamente). É máis, descoñécese o criterio de decisión estrito para a correcta clasificación, polo tanto, non é directamente extrapolable á práctica forense. Así mesmo, o sistema categorial de análise de contido non é metódico, polo tanto, non válido. Por outra banda, vítimas de ASI/ASA teñen unha probabilidade de .70 de desenvolver un trastorno internalizante e un 30% de dano en secuela xeral, depresión e ansiedade. Concretamente, a probabilidade dun dano crónico (distimia) foi significativamente superior a un dano máis severo. En conclusión, a vitimización de ASI/ASA repercute no axuste psicolóxico na idade adulta, tendo que cuantificase a través da avaliación da actividade global da vítima (EEAG).

**Palabras chave:** Hipótese Undeutsch; criterios de realidade; CBCA; pegada psicolóxica; abuso sexual infantil e adolescente.



**FORENSIC ASSESSMENT OF THE  
CREDIBILITY OF A TESTIMONY AND  
INTERNALIZING SYMPTOMATOLOGY  
IN CRIMES COMMITTED IN  
PRIVATE SPHERE**

**ABSTRACT**

Credibility assessment of victim's testimony is a key proof element in private crimes as testimony is frequently the only available evidence. The technique used worldwide to evaluate testimony reality is the Criteria Based Content Analysis (CBCA), which is funded in the *Undeutsch Hypothesis*: memories of real-life self-experienced events are different in content and quality (reality criteria) from memories of fabricated events. In spite of its extended use in justice, in scientific background contradictory apparently opinions and findings were expressed regarding its discriminative capacity between both memories and, hence, its applicability to the forensic setting. On its behalf, credibility assessment jointly with internalizing sequelae assessment as a consequence of child sexual abuse (CSA) or adolescent sexual abuse (ASA) victimization, give support to statement veracity and contribute to evidence sufficiency. According to this state of the art, validating Hypothesis through reality criteria was proposed. Likewise, technique admissibility in court will be analysed throughout *Daubert Standards* and law of precedence requirements analysis. Furthermore, injury quantification (depressive and/or anxiety disorder) and its development probability as a consequence of victimization were proposed. The results showed *Undeutsch Hypothesis* and CBCA are valid, they discern between vivid and fabricated memories, in different populations (i.e., children and adults) and contexts. Although the technique meets *Daubert Standards*, classification error rate of false statements as truthful in field studies is not admissible (9.8% vs 28.5%, in children and adults, respectively). Moreover, the strict decision criterion to classify correctly is unknown; hence, it is not directly extrapolable to forensic settings. Likewise, the content analysis categorical system is not methodic, consequently not valid. On the other hand, ASI/ASA victims have a .70 probability of developing an internalizing disorder and 30% damage quantification in general sequelae, depression and anxiety. Concretely, the probability of a chronic injury (dysthymia) was significantly superior to a severe injury. In short, ASI/ASA victimization impact psychological adjustment in adulthood, it should be quantified based on the victim's global assessment functioning (GAF).

**Keywords:** *Undeutsch Hypothesis*; reality criteria; CBCA; psychological injury; child and adolescent sexual abuse.





# ÍNDICE

<b>1. INTRODUCCIÓN</b> .....	<b>21</b>
1.1. LA PERICIAL PSICOLÓGICA Y EL ANÁLISIS DE LA CREDIBILIDAD DEL TESTIMONIO .....	21
1.2. INSTRUMENTOS DE EVALUACIÓN DE LA REALIDAD DEL TESTIMONIO .....	24
1.3. STATEMENT VALIDITY ASSESSMENT (SVA) .....	25
1.3.1. Origen y estructura .....	25
1.3.2. Admisibilidad del SVA en los distintos países.....	27
1.4. CRITERIA-BASED CONTENT ANALYSIS (CBCA) .....	28
1.4.1. Paradigmas de investigación. Estudios de campo y experimentales .....	31
1.4.2. Limitaciones de la técnica .....	32
1.5. ABUSO SEXUAL EN LA INFANCIA Y ADOLESCENCIA. SECUELAS INTERNALIZANTES DE LA VICTIMIZACIÓN.....	34
1.6. EL PROCEDIMIENTO META-ANALÍTICO .....	37
1.6.1. Tipos de meta-análisis .....	40
<b>2. OBJETIVOS</b> .....	<b>43</b>
<b>3. MÉTODO</b> .....	<b>45</b>
<b>4. RESULTADOS</b> .....	<b>49</b>
<b>5. DISCUSIÓN</b> .....	<b>55</b>
<b>6. DISCUSSION</b> .....	<b>65</b>
<b>7. REFERENCIAS</b> .....	<b>75</b>
<b>8. ARTÍCULOS DE INVESTIGACIÓN</b> .....	<b>83</b>
8.1. UNDEUTSCH HYPOTHESIS AND CRITERIA BASED CONTENT ANALYSIS: A META- ANALYTIC REVIEW.....	85
8.2. CRITERIA-BASED CONTENT ANALYSIS (CBCA) REALITY CRITERIA IN ADULTS: A META-ANALYTIC REVIEW .....	107
8.3. PSYCHOLOGICAL INJURY IN VICTIMS OF CHILD SEXUAL ABUSE: A META- ANALYTIC REVIEW.....	123
8.4. ANÁLISIS DE CONTENIDO EN DECLARACIONES DE AGRESORES. UNA REVISIÓN META-ANALÍTICA .....	153
<b>9. ÍNDICE DE TABLAS</b> .....	<b>161</b>
<b>10. APÉNDICE</b> .....	<b>173</b>
10.1. RESUMEN .....	175
10.2. ARTÍCULOS.....	185



# 1. INTRODUCCIÓN

## 1.1. LA PERICIAL PSICOLÓGICA Y EL ANÁLISIS DE LA CREDIBILIDAD DEL TESTIMONIO

En la actualidad, la Ley 1/2000 de Enjuiciamiento Civil (LEC), en sus artículos 335 a 352, regula la intervención de los peritos en el estado español. Asimismo, la Ley de Enjuiciamiento Criminal (LECrim) en su art. 456, hace mención a que el Juez podrá solicitar un informe pericial, cuando no disponga de los conocimientos científicos, técnicos o artísticos necesarios para apreciar determinadas circunstancias del caso. Sin embargo, dicho informe no será vinculante para el dictamen del juez o tribunal, sino que su interpretación, basada en la *sana crítica*, servirá de apoyo junto con otras pruebas practicadas para valorar libremente el dictamen pericial (De Luca, Navarro, & Cameriere, 2013). Así, la función del psicólogo forense como perito judicial es la de auxiliar a la justicia, asesorando a jueces y tribunales a partir de los conocimientos propios de la disciplina.

La jurisprudencia estadounidense de los años 80 no ha sido clara en el establecimiento del papel del psicólogo forense como asistente en tribunales de justicia en casos de abusos sexuales a menores. Se acepta, en determinados estados como Hawai, Montana o Minnesota, entre otros, el testimonio experto sobre la valoración de la credibilidad de la víctima en general (*reputation evidence of character*), aunque con ciertas limitaciones. Sin embargo, se restringe el papel del perito a determinadas casuísticas: emisión de un juicio sobre la credibilidad en relación con un patrón de características comportamentales de los niños abusados sexualmente; o cuando el testimonio experto mejora la credibilidad de la víctima. En el polo opuesto, ciertos estados rechazan de plano la admisibilidad del testimonio experto por entender que la credibilidad del testimonio es competencia exclusiva del juez y, por tanto, los peritos no deben concluir sobre la culpabilidad o inocencia del acusado; o cuando el testimonio no ayuda de ningún modo al jurado o bien cuando supone un perjuicio. No obstante, ante el creciente número de casos de abuso sexual infantil en la época y el conocimiento puramente profano de jueces y tribunales sobre esta problemática, el foco de atención se dirigió a la búsqueda de formas más eficaces para la verificación de los testimonios de las víctimas (Baker, 1990). Esta necesidad detectada llevó a modificar las reglas tradicionales (*Federal Rules of Evidence*), con el objetivo de admitir el testimonio experto en casos de abuso sexual infantil, tal y como se recoge en la LECrim española, es decir, cuando el perito experto posea unos conocimientos, habilidades, dilatada experiencia o educación de la que carecen los jueces y tribunales para emitir una opinión sobre determinados comportamientos en relación con el padecimiento de sintomatología clínica (Smith vs. State), e.g. aportar pruebas al jurado sobre el diagnóstico del Trastorno por Estrés Postraumático en víctimas de abuso sexual infantil. Hasta ese momento, sólo se admitía el testimonio experto cuando las pruebas fueran aceptadas generalmente por la comunidad científica, un precepto que, aunque vago, fue incluido posteriormente en los *Daubert Standards* (se hablará más detalladamente de ellos en el epígrafe siguiente). Empero, los criterios Daubert dejaron en un segundo plano este requerimiento de las reglas tradicionales

subrayando, como criterio de mayor relevancia, el uso de instrumentos validados empíricamente por la comunidad científica.

Por su parte, en el contexto alemán, a comienzos del siglo XX el psicólogo William Stern, fue el primero en acentuar la relevancia de la evaluación de la veracidad de un testimonio, centrándose en las diferencias individuales de las personas declarantes, especialmente en aquellos casos en los que se carecía de evidencias físicas para la incriminación. Sin embargo, la fiabilidad del testimonio recaía, en muchas ocasiones, en elementos indirectos que ponían en duda la credibilidad del individuo previamente al enjuiciamiento. Por ejemplo, un historial de promiscuidad sexual (Undeutsch, 1989) o la minoría de edad (Baker, 1990) a la cual se le presupone una alta propensión a la fantasía (falso testimonio), restaban credibilidad a la declaración de mujeres y menores víctimas de abusos sexuales, respectivamente. Posteriormente, la defensa de Udo Undeutsch de una pericial psicológica en un tribunal de menores marcó un punto de inflexión en la Psicología del Testimonio en general, y en la evaluación de la credibilidad en particular. Con base en esta intervención, el tribunal concluyó que el estudio del testimonio de una víctima hecho por un testigo experto fuera de la sala de justicia lograba resultados sumamente diferentes, en términos de fiabilidad y validez, a los alcanzados durante el juicio oral. Por ello, el Tribunal Supremo alemán dictaminó la necesaria intervención de un perito expert ante la falta de otras evidencias -salvo el testimonio de la víctima- para evaluar la veracidad del testimonio de menores víctimas de abuso sexual.

Recientemente, un estudio elaborado por Arce y Fariña (2015), puso de relieve la importancia del testimonio experto. Concretamente, estos autores concluyeron que la evaluación de la credibilidad del testimonio a partir de un mandato judicial, permitió dotar de valor de prueba suficiente a la declaración en el 93.3% de los casos evaluados. Asimismo, encontraron que la mayoría de los casos sobreseídos se produjeron por la ausencia de informes periciales psicológicos, tanto de credibilidad de testimonio que avalaran la realidad de la declaración como del establecimiento de la huella psíquica consecuencia del hecho delictivo sufrido por la víctima. Igualmente, Van Koppen (2007) subrayó la importancia de la asistencia del perito experto en la toma de decisiones en casos penales graves que no son considerados “rutinarios” y que constituyen, aproximadamente, un 12% del total.

La valoración de la prueba se convierte, por tanto, en el elemento esencial para la toma de decisiones judiciales, recayendo la motivación de la sentencia, en la fiabilidad y validez de la práctica de la prueba (Arce & Fariña, 2015). La fiabilidad estará condicionada por la credibilidad otorgada a la víctima o testigo e, incluso, por la consistencia interna del propio testimonio, mientras que la validez hace referencia a la relevancia de la prueba para el caso concreto. Cuando la declaración de la víctima constituye la única prueba de cargo, para que ésta sea suficiente, la Jurisprudencia española establece tres criterios jurídicos como apoyo al proceso de toma de decisión en la valoración de la prueba: *ausencia de incredibilidad subjetiva* (ausencia de motivaciones espurias para declarar), *verosimilitud* (corroboraciones periféricas objetivas que refuercen el testimonio) y *persistencia en la incriminación* (entendida como validez de la prueba, consistencia en el tiempo y ausencia de contradicciones). El estudio realizado por Seijo (2007) en el que se analizaron sentencias judiciales, reveló que las decisiones de magistrados sobre la verosimilitud de un testimonio recaían, en gran parte, en informes psicológico forenses que analizaban la credibilidad de la declaración de la víctima y/o el establecimiento del daño psíquico consecuencia de la

victimización. Estos resultados destacan la necesidad de elaborar pruebas objetivas, científicamente validadas, que estudien criterios empíricos de credibilidad, abandonando así los criterios tradicionales subjetivos de los que disponen los jueces para la estimación de la prueba (fiabilidad) en términos de realidad.

No existe un consenso respecto de la estructura del informe pericial, sin embargo, los colegios profesiones de psicología han elaborado una serie de directrices (Seijo et al., 2014) para la redacción del contenido de un informe pericial psicológico, basándose en los requerimientos legales anteriormente expuestos:

1. Introducción. Debe incluir los datos de quien emite la pericial, su número de identificación y el número de expediente, en el caso que venga requerido por mandato judicial, así como los datos del demandante/denunciante y persona demandada o denunciada. Igualmente, se incluirá el objeto del informe. Finalmente, el número y fecha de las sesiones de evaluación y una breve mención a la intervención realizada en cada una de ellas.
2. Procedimiento o metodología de evaluación. Este punto constituye una de las partes más relevantes del informe pericial, puesto que en él se explicarán las técnicas y procedimientos utilizados para la evaluación. La mayor o menor credibilidad de nuestra pericial psicológica dependerá del rigor científico y concreción de los instrumentos seleccionados, así como las técnicas de evaluación forense utilizadas. De facto, los resultados y conclusiones del informe quedarán supeditadas a la metodología de evaluación. Por ello, el uso de técnicas forenses metódicas avaladas científicamente (criterio Daubert) permitirán tomar decisiones sobre la evaluación con un elevado nivel de certeza, evitando lo máximo posible la zona de incertidumbre.
3. Resultados. Se expondrán los resultados obtenidos tras la aplicación de las pruebas pertinentes y se incluirá la interpretación de las mismas.
4. Conclusiones. Este apartado es el resultado de la integración de los datos objetivos recabados en las anteriores fases, y deberá responder al mandato formulado por el juez o por alguna de las partes, de forma clara y concisa. Cuando el objeto del informe sea la determinación de la realidad de un testimonio, las conclusiones deben realizarse en términos de probabilidades, ajustándonos a las siguientes categorías: *declaración (muy) probablemente cierta/real/creíble/de una memoria basada en hechos vividos; carente de criterios de realidad/de una memoria basada en hechos vividos; declaración o prueba inválida; e indeterminada (también puede referirse como prueba insuficiente* (Arce & Fariña, 2006).

La exhaustividad y rigurosidad durante todo el proceso de evaluación de la víctima determinará su fiabilidad y validez y, por tanto, su admisibilidad como prueba de cargo en el juicio. Por ello, resulta imprescindible la utilización de técnicas de evaluación en el contexto forense avaladas por la evidencia científica.

## 1.2. INSTRUMENTOS DE EVALUACIÓN DE LA REALIDAD DEL TESTIMONIO

Los enfoques y herramientas que se han utilizado para la evaluación de la credibilidad del testimonio han sido muy variados. Por un lado, en el paradigma fisiológico, el polígrafo es el instrumento más extendido, especialmente en Norteamérica, donde tiene sus orígenes. El objetivo es detectar cuándo un sospechoso está mintiendo a partir de cambios comportamentales difíciles de controlar de modo consciente (e.g. cambios en el nivel de conductancia de la piel o sudoración, en la presión sanguínea). Esta premisa subyace a las distintas técnicas poligráficas, las más conocidas son: Relevant-Irrelevant Test, Comparison Question Test y el Guilty Knowledge Test. A pesar de su extendido uso a nivel experimental, la utilización del polígrafo en el proceso judicial, en general, es muy reducida, circunscribiéndose a determinados casos y tribunales, por incumplimiento de ciertos criterios Daubert de admisibilidad de pruebas científicas (Griesel & Yuille, 2007). Asimismo, como limitación de la técnica poligráfica, las respuestas fisiológicas pueden distorsionarse a través del entrenamiento y manipulación en las respuestas control, mermando así la precisión del instrumento.

Por lo que respecta al paradigma de detección de engaño a través del estudio de señales no verbales, esto es, se relaciona la mentira con determinados movimientos corporales (e.g. movimientos de manos, de cabeza) y características del habla (e.g. titubeos), las investigaciones realizadas revelan la escasa fiabilidad de estas medidas (DePaulo, et al., 2003). De facto, no permiten discriminar entre engaño y rasgos de personalidad inherentes al individuo. Tan solo identifican cambios comportamentales entre individuos, en un determinado contexto y, en ciertas ocasiones, el nivel de precisión en la detección se situaría por debajo del azar. No obstante, Vrij, Edward, Roberts y Bull (2000) consideran que la combinación de técnicas verbales y no verbales en la detección del engaño, podría alcanzar una tasa de correcta clasificación (verdaderos positivos y verdaderos negativos) de aproximadamente el 81% de los casos.

Los instrumentos que evalúan la realidad del testimonio con base en indicadores verbales han sido los más utilizados y gozan de un mayor respaldo científico en la actualidad (Vrij, Akehurst, Soukara, & Bull, 2004). Concretamente, el Statement Validity Assessment (SVA) y su componente principal, el Criteria-Based Content Analysis (CBCA). El eje central del CBCA es el análisis del contenido de la declaración y constituye el instrumento de evaluación de la validez del testimonio más estudiado y con mayor aceptación en la práctica forense. Otra herramienta que evalúa aspectos verbales del testimonio es el Reality Monitoring (RM; Johnson & Raye, 1981). La teoría que subyace a este procedimiento tiene su explicación en el proceso cognitivo que guía la capacidad para diferenciar una memoria externa basada en un hecho percibido de una memoria interna, es decir, inventada o imaginada. A diferencia del CBCA, el RM se considera una herramienta para la detección del engaño, mientras que el CBCA constituye un instrumento orientado a la búsqueda de señales de veracidad (*truth bias*). La investigación experimental en la psicología aplicada al testimonio ha recurrido al RM en numerosas ocasiones, sin embargo, su uso individualizado en la práctica forense es poco frecuente, no así su aplicación combinada con otros sistemas categoriales (Arce & Fariña, 2013; Sporer, 2004), incrementando la capacidad discriminativa o fiabilidad del sistema (Vrij, 2008).

El Scientific Content Analysis (SCAN), sistema categorial de análisis de contenido del testimonio desarrollado por Smith (2001), es utilizado por las fuerzas policiales de distintos

continentes (Bogaard, Meijer, Vrij, & Merckelbach, 2016). La teoría memorística que subyace a esta herramienta es equivalente a la denominada Hipótesis Undeutsch: las memorias de un hecho experimentado difieren en contenido y calidad de aquellas basadas en la invención o fantasía. De facto, algunos de los criterios incluidos en el SCAN se corresponden con criterios de realidad del CBCA. La diferencia radica en la unidad de análisis o *target*, el SCAN evalúa el tipo de lenguaje que utilizan las personas entrevistadas que mienten en contraposición a las utilizadas por las que responden con sinceridad. No obstante, la ausencia de una base teórica que explique la capacidad discriminativa del SCAN así como los resultados contradictorios hallados en los escasos estudios experimentales publicados hasta la fecha (Bogaard et al., 2016), han derivado en una falta de apoyo empírico hacia este instrumento, por parte de la comunidad científica, como herramienta válida para la detección del engaño.

### 1.3. STATEMENT VALIDITY ASSESSMENT (SVA)

#### 1.3.1. Origen y estructura

Si bien originariamente el juicio sobre credibilidad se emitía con base en la personalidad o comportamiento de la víctima, Undeutsch desvió el foco atencional hacia las características del propio testimonio. Con ello se pretendía evitar una posición desfavorable de partida de la víctima y una evaluación puramente subjetiva basada en el sentido común. De este modo, independientemente de si la persona nos inspira confianza o no, la calidad de su declaración recaerá en las características de la propia narración. Para la evaluación su calidad, Undeutsch (1989) propone una serie de “criterios de realidad” que se esperan encontrar con mayor frecuencia en un testimonio real. Este autor se basa en la premisa de que las memorias de hechos vividos o auto-experimentados difieren significativamente de aquellas memorias falsas (inventadas, fabricadas, distorsionadas) siendo aquellas de mayor calidad en términos de criterios. A esta afirmación se le conocerá posteriormente como la Hipótesis Undeutsch (Steller, 1989) y al procedimiento criterial como Statement Reality Analysis (SRA).

Este nuevo enfoque fue aceptado por investigadores de distintas partes del mundo, y muchos de ellos han colaborado en la modificación, redefinición y actualización de la propuesta (Undeutsch 1989). El procedimiento se componía de dos partes: una entrevista con preguntas abiertas y el análisis de la declaración mediante un listado de criterios de realidad. Esta estructura y metodología fue sistematizada posteriormente por Steller y Köhnken (1989) dando lugar a lo que hoy en día se conoce como Statement Validity Assessment (SVA), herramienta de evaluación de la realidad de la declaración más ampliamente utilizada en la actualidad (Vrij, 2008). El objetivo del SVA es determinar el origen de la declaración y si ésta contiene aspectos propios de experiencias realmente vividas por la persona entrevistada, sin emitir conclusiones sobre la credibilidad general de la víctima (Raskin & Esplin, 1991). El SVA se compone de cuatro fases: 1) el estudio del caso; 2) una entrevista semi-estructurada, en la que la persona declarante, sin influencias por parte del entrevistador/a, relata el suceso vivido; 3) el análisis de contenido de la declaración a través del Criteria-Based Content Analysis (CBCA); y 4) una combinación de los resultados anteriores con información derivada de un set de preguntas denominado *Listado de Validez*.

1) Estudio de caso: previamente a la realización de la entrevista, se estudia minuciosamente la información disponible relativa al caso: características de la víctima (e.g., edad, desarrollo cognitivo), declaraciones previas (con qué frecuencia ha sido entrevistada la víctima, ha habido o no inconsistencias entre las declaraciones) y otras características relacionadas con las motivaciones para testimoniar. Todo ello servirá para analizar el origen de la declaración y los posibles motivos para declarar falsamente.

2) Entrevista semi-estructurada en la que el relator o relatora narre libremente los hechos acaecidos (a partir de técnicas de entrevista que favorezcan el relato libre: Entrevista Cognitiva, Fisher & Geiselman (1992); Step-Wise Interview Guidelines, Yuille, Marxsen, & Cooper, 1999). La pericia del entrevistador/a es un elemento crucial en la aplicación del SVA, ya que el modo en que se manejan los silencios, la forma en que se realicen las preguntas y las actitudes para con el entrevistado, pueden condicionar sus respuestas. Así, el procedimiento comenzaría con una narración libre de los hechos, seguida de una serie de preguntas abiertas, no dirigidas ni sugestivas, evitando así contaminar la declaración. Esto resulta de vital importancia en el caso de víctimas menores edad, ya que el nivel de sugestionabilidad es superior al de los adultos. La información de caso conjuntamente con los datos aportados por el relator durante la entrevista, son de especial relevancia para el proceso de falsación de hipótesis (Popper, 1968, citado en Köhnken, 2004), es decir, se trata de someter a evaluación las hipótesis alternativas a la hipótesis principal de partida (i.e., la validez de la declaración), procedimiento análogo al diagnóstico diferencial en el ámbito clínico. Las hipótesis alternativas (Köhnken, 2004; Raskin & Esplin, 1991) sometidas a contraste serían: 1) la declaración es completa e intencionalmente falsa, es decir, nunca le ha ocurrido a la persona declarante, con motivaciones de venganza o con la finalidad de ayudar a otra persona; 2) la mayor parte de la declaración es válida, sin embargo, contiene elementos relevantes que han sido inventados como, por ejemplo, a consecuencia de presiones; 3) la declaración es válida, no obstante, se acusa a un individuo distinto al que realmente cometió el delito (transferencia inconsciente); 4) la declaración es un testimonio "prestado", es decir, no ha sido vivenciado por quien declara pero describe lo que le ha ocurrido a otras personas, como por ejemplo, en los casos de custodia durante un proceso de separación o divorcio; 5) el testimonio es fruto de la fantasía, como consecuencia de una enfermedad mental que le impide discernir entre realidad y ficción; y 6) la declaración es falsa, pero no de modo intencional, sino que puede tratarse de una falsa memoria del evento. El proceso de falsación de hipótesis alternativas comienza en una fase anterior a la entrevista, esto es, durante el análisis minucioso de caso (Köhnken, 2004). En esta fase se analiza la información relativa a la víctima o testigo, que es relevante para el caso y que podría explicar el origen de la narración (i.e., edad, habilidades cognitivas, hecho aislado o repetido en el tiempo, número de entrevistas a las que han sido sometidos la víctima o testigo, inconsistencias entre las declaraciones, etc.). Esta información, que será recuperada de nuevo en el análisis de la lista de validez, es determinante para la evaluación de la declaración como un todo. Finalmente, la totalidad de la entrevista deberá ser grabada en vídeo para proceder, posteriormente, a su transcripción y codificación.

3) Una vez transcrita la entrevista, se aplicará el CBCA con el objetivo de evaluar la realidad de la declaración, a través del análisis de su contenido. Se abordará de forma más pormenorizada el instrumento en un epígrafe posterior.



4) Finalmente, se aplica la Lista de Validez con el fin de conocer la plausibilidad de las hipótesis alternativas planteadas, a partir de un análisis holístico de la información disponible (estudio de caso), de los datos recopilados en la entrevista y del análisis de contenido del testimonio. Llegados a este punto, estaremos en disposición de tomar una decisión razonada sobre la realidad de la declaración, ya que la puntuación obtenida en el CBCA no es suficiente para concluir sobre la veracidad del testimonio (Vrij, 2008). Esta Lista de Validez está conformada por cuatro categorías generales de información que deben ser evaluadas: características psicológicas (e.g. capacidades cognitivas del declarante, susceptibilidad a la sugestión), características relativas a la entrevista (e.g. malas prácticas en el proceso de entrevista), motivaciones para declarar (i.e., presiones para hacer una falsa denuncia, algún tipo de relación entre la víctima y el acusado), otras cuestiones relativas a la investigación que pueden arrojar dudas sobre la validez de la declaración (e.g. descripción de hechos contraria a las leyes de la naturaleza, contradicciones intra e inter-declaración). Cuando la declaración es válida, esto es, de una alta calidad, se utiliza la *checklist* para desechar las hipótesis alternativas. Si la declaración es de baja calidad, el objetivo del listado es conocer si la información adicional apoya alguna de las hipótesis alternativas, o si la baja calidad de la declaración se debe a una pobre entrevista y/o a una limitada capacidad cognitiva de la víctima (Raskin & Esplin, 1991).

### 1.3.2. Admisibilidad del SVA en los distintos países

La necesidad de evaluar la calidad de un testimonio, en términos de credibilidad, surge a mediados del siglo XX, cuando se puso de manifiesto la labor del perito como experto entrevistador. El perito poseía la capacidad de obtener mayor información y menos sesgada en comparación con la información obtenida en el contexto del juicio oral. El uso del CBCA como herramienta de evaluación de la realidad de un testimonio basado en criterios específicos de contenido es, hasta nuestro tiempo, el único método con apoyo científico aceptado en los tribunales alemanes. Especialmente, cuando se trata de delitos contra la libertad e indemnidad sexuales, ya sea en víctimas adultas o menores de edad (Steller & Böhm, 2006).

El panorama en EEUU, como ya se ha indicado, era cualitativamente distinto. La admisibilidad de una prueba pericial de carácter científico requería del cumplimiento de los criterios Daubert o factores de científicidad (Vázquez-Rojas, 2014): 1) ¿Puede probarse la hipótesis científica? 2) ¿Ha sido probada la hipótesis? 3) ¿Se conoce la tasa de error de la técnica? 4) ¿Ha sido sometida la hipótesis o la técnica a un proceso de revisión por pares y publicación? 5) ¿Se basa la hipótesis y/o técnica en una teoría generalmente aceptada como válida por la comunidad científica? A pesar del establecimiento de esta norma, seguía existiendo una desconfianza hacia la capacidad de los expertos para determinar la calidad de una declaración de abuso sexual con una alta fiabilidad (McGough, 1991) descartando, por tanto, la apreciación de la prueba en la sala de justicia.

Los criterios Daubert surgen en el contexto judicial americano a raíz del caso *Daubert vs. Merrel Dow Pharmaceuticals* (1993). El Tribunal Supremo americano, con el objetivo de no caer en la perspectiva reduccionista de asimilar la científicidad de la prueba con conocimiento fiable *per se*, dictaminó una serie de pautas que toda prueba pericial debe cumplir para ser

admitida en un proceso judicial. Este enfoque es defendido en el sistema de justicia español como una forma de introducción de criterios técnicos de valoración judicial de la prueba garantizando así la tutela judicial efectiva (Manzanero y Muñoz, 2011). En consecuencia, este requerimiento legal obliga al uso de herramientas en la justicia que cumplan con los criterios Daubert. En tanto que la fiabilidad de la prueba recae en la evaluación de criterios de credibilidad de carácter empírico (Arce, Seijo, & Novo, 2009), que se encuentran perfilados en la técnica SVA, ésta se erige como una alternativa robusta para el análisis del testimonio. Los criterios de realidad constituyen criterios empíricos, objetivos y con respaldo científico, mientras que los legales tienen su base en indicadores subjetivos para la estimación de la credibilidad de la declaración (Arce & Fariña, 2015).

#### **1.4. CRITERIA-BASED CONTENT ANALYSIS (CBCA)**

El CBCA es el componente del SVA que más se ha estudiado, por ser el principal y el de mayor utilidad empírica (Vrij, 2008). Contrariamente, el Listado de Validez ha suscitado menos investigación como consecuencia de las escasas investigaciones y por la dificultad de análisis del procedimiento de entrevista (Horowitz, 1991). En un intento por sistematizar el procedimiento y elaborar definiciones más precisas de los criterios, Steller y Köhnken (1989) propusieron un listado de 19 criterios de realidad organizados en 5 grandes categorías: características generales (criterios 1-3), contenidos específicos (criterios 4-13), contenidos relacionados con la motivación (criterios 14-18) y elementos específicos de la ofensa (criterio 19).

**Características generales** (se refieren a la declaración en su conjunto).

*Criterio 1: Estructura lógica.*

El testimonio goza de consistencia interna y coherencia contextual.

*Criterio 2: Elaboración inestructurada.*

Los elementos que conforman el hecho delictivo se encuentran desorganizados en el relato, no siguiendo un orden cronológico de los acontecimientos. Sin embargo, esos elementos unificados forman un todo coherente.

*Criterio 3: Cantidad de detalles.*

El testimonio se compone de numerosos y diferentes detalles.

**Contenidos específicos** (partes específicas del testimonio que le aportan viveza).

*Criterio 4: Engranaje contextual.*

Ubicación de los hechos en un espacio y en un tiempo.

*Criterio 5: Descripción de interacciones.*

Cadena de acciones y reacciones que se suceden durante el incidente, entre víctima/testigo y otros actores implicados.

*Criterio 6: Reproducción de conversaciones.*

Descripción literal de conversaciones ocurridas entre víctima/testigo y otros actores implicados. Los hablantes deben ser fácilmente reconocibles y los diálogos una réplica de las palabras utilizadas por alguno de los hablantes.

*Criterio 7: Complicaciones inesperadas durante el incidente.*

Sucesos inesperados que cortan la trayectoria del evento como, por ejemplo, una interrupción imprevista.

**Peculiaridades del contenido** (partes específicas del testimonio que le aportan viveza).

*Criterio 8: Detalles inusuales.*

Detalles que, por su naturaleza extraña aunque no irreal, es poco probable que ocurran.

*Criterio 9: Detalles superfluos.*

Detalles irrelevantes descritos en conexión con el hecho relatado, pero que no contribuyen significativamente a los hechos.

*Criterio 10: Incomprensión de detalles relatados con precisión.*

Interpretación incorrecta de un detalle correctamente descrito. Este criterio está directamente conectado con el desarrollo cognitivo del niño, describiendo aspectos de forma correcta pero que por su inmadurez cognitiva no alcanza a comprender.

*Criterio 11: Asociaciones externas relacionadas.*

Consiste en hacer referencia a un evento externo a los hechos, aunque relacionado con la naturaleza de éstos. Por ejemplo, en un caso de abuso sexual, hacer mención a conversaciones pasadas de naturaleza sexual entre víctima y agresor.

*Criterio 12: Relatos del estado mental subjetivo.*

Referencias a sentimientos, emociones y cogniciones propias.

*Criterio 13: Atribución del estado mental del autor del delito.*

Referencias a los estados mentales y motivos del agresor, así como descripción de reacciones afectivas y estados fisiológicos.

**Contenidos referentes a la motivación** (elementos que se relacionan social y estereotípicamente con la mentira. Estos criterios suelen evitarse al creer que de ese modo dan una imagen de honestidad. También acuñados por Ruby y Brigham (1998, citado en Vrij, 2005) como *contrary-to-truthfulness-stereotype*.

*Criterio 14: Correcciones espontáneas.*

Incluir cambios o correcciones en la declaración de forma espontánea, en un intento de mejorar el testimonio.

*Criterio 15: Admisión de falta de memoria.*

Admisión de lagunas de memoria.

*Criterio 16: Plantear dudas sobre el propio testimonio.*

Expresar dudas, en parte o en la totalidad del testimonio, sobre las propias correcciones hechas durante la declaración.

*Criterio 17: Auto-desaprobación.*

Referencias a detalles o comportamientos autoincriminatorios, actitud crítica sobre la propia conducta, en relación con el abuso.

*Criterio 18: Perdón al autor del delito.*

La declaración favorece al agresor, a través de explicaciones o exoneraciones de su conducta.

**Elementos específicos de la ofensa** (características criminógenas definitorias del delito, en muchas ocasiones contrarias al sentido común).

*Criterio 19: Detalles característicos de la ofensa.*

Detalles de la situación de abuso contrarios al saber común, contradicen las creencias habituales sobre un delito de esta naturaleza, solamente identificables por profesionales con conocimientos profundos en esta materia.

Los primeros trece criterios se corresponden con el componente cognitivo de la hipótesis, esto es, solo una persona con unas determinadas habilidades cognitivas y verbales, que haya experimentado el evento narrado, incluirá con mayor probabilidad estos criterios en su declaración. Por otro lado, la hipótesis también contempla un componente motivacional (criterios 14-18), un individuo que deliberadamente fabrica una historia evitará incluir en su relato cualquier característica (criterio de realidad) que, por sentido común, cree que le hará parecer poco honesto (*impression-management*) (Köhnken & Steller, 1988).

El análisis de contenido se realiza sobre la transcripción de la entrevista, tomando como base los criterios del CBCA. Se codificará la ausencia o presencia de cada criterio en la declaración, o en qué medida el criterio está presente (ausente-presente-fuertemente presente). Para ello, al menos dos evaluadores independientes deberán codificar la calidad del testimonio en términos de criterios de realidad. Posteriormente, se calculará la concordancia o fiabilidad inter-jueces, entendida ésta como la exacta correspondencia entre los evaluadores en la identificación de cada criterio de realidad, es decir, ambos codificadores deben coincidir no sólo en el criterio identificado sino en el lugar exacto de la transcripción en el que fue identificado. Es habitual en los estudios experimentales que ponen a prueba la validez del CBCA, medir la fiabilidad del instrumento a partir del porcentaje de acuerdo entre evaluadores, sin embargo, esta medida no sería adecuada (Horowitz, 1991) puesto que no tiene en cuenta las variaciones debidas al azar, la estabilidad temporal de la fiabilidad (medida intra-jueces) y el error de medida. La mayor fuente de variabilidad observada en las codificaciones fue hallada en los criterios individuales (Horowitz et al., 1997) no alcanzando, además, algunos de ellos, una fiabilidad adecuada. Por el contrario, la revisión cualitativa de Vrij (2008) puso de manifiesto el alto índice de fiabilidad obtenido en la puntuación total del CBCA. Estos resultados llevaron a algunos autores a redefinir (Horowitz et al., 1997) o unificar (Roma, San Martini, Sabatello, Tatarelli, & Ferracuti, 2011) determinados criterios de realidad ya que no eran mutuamente excluyentes (e.g., con frecuencia los criterios “detalles

inusuales” y “detalles superfluos” se solapan, dificultando su discriminación y, afectando en último término, a su fiabilidad).

Una vez codificado el contenido de la declaración, podremos tomar una decisión sobre la calidad del testimonio, no así de su credibilidad, puesto que no se puede reducir ésta al CBCA, sino que es necesario tener en cuenta otra información relativa al caso concreto (i.e., documentación judicial del caso, entrevista con la víctima, listado de validez de información adicional de caso) (Steller & Boychuk, 1992).

En un intento por mejorar el sistema criterial de análisis de contenido, han surgido una serie de criterios adicionales complementarios al listado original del CBCA. Köhnken, Schimossek, Ascherman, y Höfer (1995), pusieron a prueba la capacidad discriminativa entre declaraciones reales y fabricadas de cinco criterios adicionales, utilizando un paradigma experimental de vídeo. La presencia en la declaración de los tres primeros (i.e., estilo de la declaración, dar explicaciones de la falta de memoria y dar muestras de inseguridad) se relaciona con señales de realidad, mientras que la presencia de los dos últimos (i.e., clichés y repeticiones) es indicativo de un testimonio probablemente falso. Asimismo, Ternes (2009) utilizó la versión alternativa de 24 criterios, 19 de ellos pertenecientes al CBCA y 5 adicionales (ver Tabla 1), para el análisis de contenido de declaraciones de agresores.

**Tabla 1. Criterios Adicionales**

<ul style="list-style-type: none"> <li>• Estilo de la declaración: se considera extensa y densa cuando el entrevistado describe aspectos irrelevantes sobre los que no se le ha preguntado (<i>long-winded</i>). Por el contrario, cuando la mayor parte del contenido del testimonio se refiere a detalles importantes y centrales del episodio (<i>issue related style</i>), es un indicador de veracidad.</li> <li>• Dar muestras de inseguridad: mostrar incertidumbre sobre la precisión en la descripción de un aspecto del evento.</li> <li>• Dar explicaciones de la falta de memoria: da razones por las que no es capaz de proporcionar una descripción detallada de los hechos. Las razones deben tener un origen interno, por ejemplo, “no presté atención a eso”.</li> <li>• Clichés: expresiones, palabras o pausas que introducen demoras en la declaración y que no son característicos del estilo discursivo del entrevistado.</li> <li>• Repeticiones: elementos ya descritos se repiten sin añadir nueva información.</li> </ul>
---

La integración de otros criterios adicionales, descritos con exactitud y sometidos a contrastación empírica, al listado original del CBCA, facilita la labor de discriminación entre declaraciones de hechos realmente vividos y hechos inventados. El objetivo último que se persigue es reducir a cero el número de falsos positivos (clasificar como real un testimonio de abuso sexual cuando es falso); debido a las consecuencias altamente estigmatizantes que esto conlleva para el denunciado pudiendo culminar, en el peor de los casos, con su ingreso en prisión (Steller & Köhnken, 1989).

#### **1.4.1. Paradigmas de investigación. Estudios de campo y experimentales**

Las investigaciones experimentales que ponen a prueba la capacidad discriminativa del CBCA, surgen ante la carencia de estudios con muestras reales de menores víctimas de abusos sexuales, constituyendo una alternativa para el estudio y mejora del procedimiento. El punto fuerte de las investigaciones de laboratorio es su validez interna, permitiendo un elevado

control de las variables que intervienen en el contexto de la investigación, así como el conocimiento tácito de la *ground truth*, es decir, se dispone de una alta confiabilidad sobre la veracidad o falsedad de la declaración. Por el contrario, su validez ecológica o de generalización de resultados es deficiente, al no poder equipararse el contexto de investigación con el contexto real de victimización (e.g., abuso sexual, violencia de género).

Para incrementar la validez ecológica de la investigación experimental en el uso del CBCA, Steller (1989) propuso tres características que los estudios de laboratorio deben reunir para que el contexto sea equiparable al contexto real de victimización de abuso sexual. Los requisitos de carácter psicológico consistirían en: 1) una implicación directa de la víctima en el suceso relatado, 2) que dicho suceso conlleve una pérdida de control durante el transcurso de los hechos, 3) y que constituya un evento emocionalmente negativo para la víctima. Adicionalmente, Volbert y Steller (2014), resaltaron otras variables que favorecen la equiparación de la demanda cognitiva y las motivaciones para declarar entre una situación real y un contexto de laboratorio: realizar entrevistas cara a cara donde se oculte que se está engañando, que los testimonios sean suficientes en longitud, seguir el procedimiento de relato libre y preguntas cuando sea necesario, o ser entrevistado en varias ocasiones. Otra de las limitaciones de las investigaciones experimentales radica en la falta de información contextual en la que se articula la declaración, haciendo imposible la contrastación de hipótesis alternativas a la realidad de los hechos (falsación de hipótesis).

Por otra parte, los estudios de campo adolecen de validez interna ante la imposibilidad de controlar las variables intervinientes y las dificultades para establecer la *ground truth* apoyándose, habitualmente, en sentencias judiciales que condenen al agresor; en pruebas o evidencias médicas; en el reconocimiento de culpabilidad del acusado; etc. Sin embargo, a diferencia de las investigaciones experimentales, la capacidad de generalización de los resultados obtenidos en estudios con muestras reales es mucho mayor.

#### **1.4.2. Limitaciones de la técnica**

La falta de consenso entre la comunidad científica respecto de la aplicabilidad del CBCA recae, por un lado, en las limitaciones que se circunscriben a la propia técnica y, por otro, a la mala praxis del personal que realiza la entrevista y la falta de entrenamiento o experiencia reglada de quien codifica la transcripción. En primer lugar, abordaremos aquellas cuestiones inherentes al CBCA que lo convierten en una herramienta semi-objetiva (Arce & Fariña, 2005) y, a continuación, se expondrán las exigencias para la aplicación del procedimiento, la presunción de habilidades y capacidades de quien entrevista, que orientan el éxito o fracaso del instrumento; así como determinadas características asociadas a los hechos y a las capacidades de la víctima o testigo.

En primer lugar, es importante recalcar que el CBCA no es un test psicométrico ni una herramienta de diagnóstico, sino que se trata de un procedimiento semi-estandarizado o de psicodiagnóstico (Steller & Köhnken, 1989), de carácter complejo en su aplicación e interpretación, cuyo objetivo es dotar al evaluador/a de información para emitir un juicio sobre la calidad de la declaración de la víctima. Puesto que no se trata de un test estandarizado, no existe un punto de corte (*cut-off score*) o criterio de decisión estricto (reglas de decisión de combinación de criterios) a partir del cual concluir la realidad de la declaración

con una probabilidad determinada. En relación con esto último, los criterios de realidad deben estar escrupulosamente descritos para que sean fácilmente identificados por los codificadores y, por consiguiente, que no exista solapamiento entre ellos (i.e., deben ser mutuamente excluyentes). Otra limitación es la ausencia de un manual que unifique la aplicabilidad de cada criterio y su definición según el contexto, provocando una merma en la fiabilidad inter-jueces en la evaluación de los criterios de realidad. Además, la imposibilidad de establecer un peso relativo para cada criterio (en relación con su capacidad para discriminar entre declaraciones auto-experimentadas e inventadas) facilitaría la toma de decisiones sobre la calidad de la declaración y, consecuentemente, sobre la veracidad de la misma. Finalmente, no se trata de un instrumento de detección de engaño, sino de búsqueda de señales de realidad (truth bias) lo que impide tomar decisiones sobre si la persona entrevistada puede estar engañando a quien evalúa (Griesel & Yuille, 2007).

Otra variable de carácter personal o individual que puede afectar a la validez de la técnica es la edad de la persona entrevistada: a mayor edad, mayor número de criterios de realidad contendrá el testimonio, mayor será la puntuación total en el CBCA (Vrij, 2008), y mayor probabilidad de clasificar la declaración como real. De igual modo, las habilidades cognitivas influyen en calidad del testimonio: a mayor capacidad cognitiva y narrativa, más detallada y mayor número de criterios de realidad contendrá la declaración. Por ende, se espera que las estrategias de engaño utilizadas sean más elaboradas cuanto más se avance en edad. De las afirmaciones anteriores se extrae la posición de “desventaja” de la que parten los menores, puesto que la minoría de edad se vincula con la inmadurez cognitiva, con dificultades para expresarse y escaso dominio del lenguaje, así como con la falta de estrategias para la recuperación de información almacenada en la memoria (Volbert & Steller, 2014). En definitiva, concluiríamos que, a menor edad, menor calidad de la declaración y, por consiguiente, menos criterios de realidad. No obstante, una investigación en las que se estudió la capacidad imaginativa de los menores (Schelleman-Offermans & Merckelbach, 2010), concluyó que aquellos que tienen una alta propensión a la fantasía elaboran relatos más detallados y, por tanto, con mayor probabilidad de contener más criterios de realidad del CBCA.

Adicionalmente, Volbert y Steller (2014) advirtieron de la influencia de una serie de variables situacionales en la calidad del relato. Cuanto menos complejos sean los hechos, mayor probabilidad de obtener un testimonio con baja calidad en términos de criterios de realidad. En otras palabras, la descripción de un hecho auto-experimentado simple en su naturaleza, podría dar lugar a un testimonio carente de detalles (Horowitz, 1991) y de extensión insuficiente y, por tanto, no podría ser sometido a un análisis de contenido (Köhnken, 2004; Steller & Köhnken, 1989). De facto, un testimonio fabricado (condición de engaño), ayudado o no para su invención con material visual, puede ser más extenso y rico en detalles que los hechos vividos pero que carecen de contenido por su simpleza (condición de realidad). A su vez, el lapso de tiempo transcurrido entre los hechos y la toma de declaración, afectan a su calidad y exactitud del recuerdo (curva del olvido; información post-suceso). Por ello, cobra especial relevancia el tipo de entrevista utilizada. Ésta debe fomentar el relato libre y facilitar el recuerdo a través de técnicas de reinstauración de contextos, evitando el uso de preguntas dirigidas y sugestivas que contaminen la memoria de los hechos y, por consiguiente, lleven a la invalidación de la declaración. Del mismo modo, con la técnica de narrativa libre evitamos la incorporación de información post-suceso al relato, que ponga en duda la persistencia de las declaraciones en el tiempo. Esta inconsistencia entre declaraciones

ocurre con frecuencia como consecuencia de los interrogatorios judiciales, especialmente en delitos contra la libertad e indemnidad sexuales y de violencia de género (Arce & Fariña, 2007).

El entrenamiento en la técnica del CBCA también contribuye a la mejora en su manejo. De facto, el meta-análisis de Hauch, Sporer, Michael, y Meissner (2014) concluyó que el entrenamiento del personal evaluador en señales verbales que indican veracidad y, especialmente, indicadores de contenido de la declaración (en contraposición a señales no verbales de engaño), incrementa las tasas de acierto, esto es, de clasificar correctamente una declaración real, con un tamaño del efecto moderado ( $g_u = 0.733$ ). Del mismo modo, para lograr una declaración suficiente y válida en términos de análisis de contenido, el testimonio debe obtenerse siguiendo los estándares del SVA (Steller, 1989).

### **1.5. ABUSO SEXUAL EN LA INFANCIA Y ADOLESCENCIA. SECUELAS INTERNALIZANTES DE LA VICTIMIZACIÓN**

Si bien el CBCA, y por consiguiente el SVA en su conjunto, se erige como una técnica robusta para el análisis de la credibilidad del testimonio, en ocasiones resulta insuficiente como prueba de cargo para tomar una decisión sobre la victimización de los hechos. En este sentido, para la corroboración del testimonio, éste debe sustentarse, no solo en el análisis de su credibilidad sino también en la acreditación de su veracidad mediante la corroboración de circunstancias externas o periféricas (verosimilitud), esto es, el establecimiento de la relación causa-efecto entre el delito y las posibles lesiones (i.e., psicológicas) consecuencia de la victimización. Por ello, resulta de especial relevancia la medida del daño en casos donde el agresor reconoce la relación sexual pero niega que no haya sido consentida por la víctima (Echeburúa, Corral, & Amor, 2002).

En respuesta a estas limitaciones prácticas, Arce y Fariña (2006), crearon y validaron una técnica de evaluación psicológico-forense que va más allá del análisis de la realidad del testimonio, denominado Sistema de Evaluación Global (SEG). Esta técnica es el resultado de la compilación y sistematización de diferentes instrumentos y técnicas de evaluación de la credibilidad del testimonio, utilizando la *Entrevista Cognitiva Mejorada* (Fisher & Geiselman, 1992) para obtener la declaración sobre los hechos en el marco del SVA; del estudio del daño psíquico consecuencia de la victimización, obteniéndose información del estado psicológico a través de la *Entrevista Clínico-Forense* (Arce & Fariña, 2001), y de la capacidad para prestar testimonio (evaluación de las capacidades cognitivas). El estudio de las propiedades psicométricas de la entrevista clínico-forense (Vilariño, Arce, & Fariña, 2013) mostró una alta fiabilidad tanto para los criterios diagnósticos del daño psicológico ( $\alpha = .850$ ) como para las estrategias de simulación ( $\alpha = .744$ ). Asimismo, corroboraron su validez predictiva (el diagnóstico del daño es similar al esperado), convergente en daño (el diagnóstico de daño en la entrevista correlaciona con la evaluación psicométrica de daño) y discriminante para las estrategias de simulación (las víctimas reales no fueron clasificadas como simuladoras).

La evaluación de la lesión o daño psíquico se obtiene a través de la medida de los efectos del delito en la salud mental de la víctima, especialmente sintomatología internalizante (e.g. depresión, ansiedad), debiéndose descartar la simulación de síntomas (DSM-V; American



Psychiatric Association, 2013). El estudio de la huella psíquica y su identificación con el delito sufrido permiten establecer la relación causa-efecto entre ambos y, consecuentemente, respaldar y reforzar la veracidad del testimonio de la víctima. El Trastorno por Estrés Postraumático (TEP) se identifica con la huella psíquica propia de delitos como agresiones sexuales o violencia en general. Este trastorno raramente se da de forma aislada, de hecho, entre el 50% y el 60% de los casos en los que se diagnostica TEP, la depresión aparece como trastorno comórbido (Blanchard et al., 2004; O'Donnell, Creamer, & Pattison, 2004). Cuando la declaración es válida y fiable y cuando se determina la relación causa-efecto entre el delito y la huella psicológica, se puede concluir que la víctima presenta una secuela consecuencia de la victimización de los hechos relatados. En el estudio de la huella es crucial descartar otros posibles eventos vitales estresantes como causa de la sintomatología relatada por la víctima.

La sintomatología o lesión psicológica asociada a la victimización de delitos de abuso o agresión sexual en la infancia es numerosa y variada (Echeburúa, Corral, Zubizarreta, & Sarasúa, 1995): ansiedad, conductas fóbicas y de evitación, baja autoestima, depresión o distimia, entre otras. Esta sintomatología constituye una medida indirecta del TEP, y puede actuar como potenciadora del diagnóstico de este trastorno (Arce & Fariña, 2005). Sin embargo, estos síntomas tomados de forma aislada no pueden justificarse como prueba de cargo, pues es posible que sean consecuencia de otras vivencias independientes del hecho delictivo (Arce & Fariña, 2005). Es por esto que el TEP se toma como la huella psíquica directamente relacionada con hechos traumáticos, permitiendo el establecimiento de la relación causa-efecto. Por otro lado, el daño puede expresarse a través de sintomatología externalizante que, en muchas ocasiones, se confunde con conductas propias de la vivencia del proceso judicial en el que se encuentra inmersa la víctima.

El abuso sexual infantil (ASI) se define como la involucración de un menor en actividades sexuales que no alcanza a comprender o para las cuales está evolutivamente inmaduro y, por tanto, no tiene capacidad para dar su consentimiento expreso (World Health Organization, WHO, 1999). Estudios epidemiológicos y meta-analíticos han puesto de relieve el problema de salud pública que constituyen los abusos sexuales en la infancia y adolescencia, con una alta prevalencia a nivel mundial (Pereda, Guilera, Forns, & Gómez-Benito, 2009; Stoltenborgh, van Ijzendoorn, Euser, & Bakermans-Kranenburg, 2011). En un intento por acercar posturas a la concepción de minoría de edad cuando nos referimos a abusos sexuales a menores, determinados autores han diferenciado el ASI (< 14 años) del abuso sexual en la adolescencia (ASA, entre 14 y 18 años), identificando dos tramos diferenciados (*cut-off*). Aunque no hay acuerdo para definir el período de la infancia y la adolescencia, habitualmente se toman los 18 años, o mayoría de edad legal, como edad tope para englobar ASI y ASA.

Las secuelas psicológicas, i.e., depresión y ansiedad, consecuencia de ASI/ASA, no solo se manifiestan a corto plazo, sino que, en muchas ocasiones, afectan al funcionamiento en la edad adulta (consecuencias a largo plazo), informando de un desajuste emocional en la adultez, variable en severidad, y que puede llegar a cronificarse (Hillberg, Hamilton-Giachritsis, & Dixon, 2011). Diversos estudios han constatado que víctimas de ASI/ASA presentan una mayor probabilidad de sufrir sintomatología internalizante en la adultez en comparación con grupos de no víctimas (Fergusson, McLeod, & Hordwood, 2013; Pérez-Fuentes, et al., 2013; Thompson et al., 2003). Para la evaluación de las lesiones en la salud mental durante la etapa adulta, investigaciones sobre las secuelas a largo plazo del ASI/ASA

se han valido de métodos de investigación retrospectivos. Estos procedimientos no están exentos de limitaciones (Briere, 1992): dificultades memorísticas para reproducir los abusos sufridos (e.g. la incorporación de información post-suceso que pueda afectar a la exactitud de los recuerdos) así como las dificultades para establecer una relación causal entre la victimización de ASI/ASA y las secuelas psicológicas en la adultez. Sin embargo, medidas de naturaleza retrospectiva tienden a infra-estimar la cifra de abuso sexual (falsos negativos) más que a sobre-estimar, siendo insuficiente este sesgo para invalidar los diseños retrospectivos de casos-contróles (Hardt & Rutter, 2004) ya que, en muchas ocasiones, es la única forma de evaluar el abuso sexual. Una aproximación multimétodo (Arce, Fariña, & Vilariño, 2015; Fariña, Arce, Vilariño, & Novo, 2014) al caso concreto de abuso sexual, esto es, la combinación de la entrevista clínico-forense que favorece la aparición espontánea de sintomatología relacionada con la situación traumática, y el uso de instrumentos psicométricos válidos y fiables, que incorporen escalas de validez para el control del engaño, disminuyen el sesgo de la medida retrospectiva.

Cuando la víctima tiene menos de 16 años, la victimización de actos de carácter sexual está considerada, en todo caso, un delito (siempre y cuando no se trate de relaciones consentidas con otro u otra menor con similar edad o desarrollo cognitivo). En estos casos, el diagnóstico de TEP para el establecimiento de la relación causa-efecto entre los hechos y las consecuencias psicológicas del delito es secundario, cobrando mayor relevancia la cuantificación de las secuelas clínicas, i.e, depresión y ansiedad, como medida indirecta de la victimización. Puesto que en el ámbito forense, se debe informar de todas las alteraciones psicológicas que presenta la víctima con el fin de elaborar un perfil de daño moral y evaluar/cuantificar las consecuencias psicológicas consecuencia de los abusos, la identificación de un cuadro clínico de depresión y/o ansiedad en la víctima, constituye un indicador clave de victimización complementario a la medida del TEP y, por tanto, de detección de huella (Arce & Fariña, 2015).

Cuando el daño psicológico sufrido en la infancia o adolescencia se prolonga hasta la edad adulta y afecta a diversas esferas de la vida de las víctimas (McKnight & Kashdan, 2009; McKnight, Monfort, Kashdan, Blalock, & Calton, 2016), debe ser operativizado y expresado en el informe pericial. Por ello, la técnica de Arce y Fariña (2009) de evaluación del daño moral en diferentes casuísticas (e.g., accidentes de tráfico), contempla la cuantificación, medida en porcentajes, de la lesión o huella psíquica consecuencia de la victimización. Para ello, se valen de la Escala de Evaluación de la Actividad Global (EEAG, eje V DSM-IV-TR) para evaluar el funcionamiento de la víctima. Todo ello, en el marco del contexto forense, donde debemos sospechar simulación de síntomas como resultado de una motivación espuria dentro del proceso penal (e.g., por cuestiones de venganza). En la última edición del DSM (2013) se eliminó la escala EEAG por considerarla subjetiva y se propuso, en su lugar, el uso del WHODAS 2.0 (World Health Organization Disability Assessment Schedule 2.0), un instrumento clínico, para cuantificar el nivel de incapacidad de la persona sometida a evaluación propuesto por la Organización Mundial de la Salud. Sin embargo, su carácter auto-administrado y su deficiencia para la evaluación de la simulación, lo hacen ineficaz para su uso en el contexto forense. La recomendación de diversos autores es continuar con el uso de la escala EEAG (Gold, 2014), que se había mostrado válida en sus dimensiones (Pedersen & Karterud, 2012). En definitiva, la cuantificación del daño en términos de desajuste psicológico en diferentes escenarios de la vida diaria, no solamente tiene implicaciones a nivel clínico (prevención e intervención), sino también a nivel legal.

## 1.6. EL PROCEDIMIENTO META-ANALÍTICO

Hasta mediados del siglo pasado, la técnica utilizada para integrar resultados derivados de investigaciones primarias era la revisión narrativa o subjetiva. Su objetivo consistía en llegar a conclusiones generalizables, mediante la acumulación e interpretación de datos cualitativos. Sin embargo, su carácter asistemático y su falta de objetividad en el proceso de selección de estudios, las tornaron poco atractivas para su publicación en revistas científicas.

Ante la falta de rigurosidad de las revisiones narrativas (la selección de los estudios se hace de forma subjetiva, al igual que la asignación de los pesos a cada artículo original), junto a otras limitaciones anteriormente mencionadas, comenzaron a surgir otros métodos de combinación de resultados que recibieron el nombre genérico de meta-análisis (Glass, 1976). Este autor definió el meta-análisis como el análisis estadístico de los resultados producto de una larga cantidad de investigaciones individuales con el propósito de integrar los hallazgos de todas ellas. Esta definición deja entrever los tres pilares definitorios del método meta-analítico: 1) integrar cuantitativamente los resultados obtenidos en investigaciones primarias sobre una misma problemática; 2) con el propósito de acumular conocimiento científico, de un modo objetivo y sistemático, sobre la temática bajo estudio; 3) a través de la aplicación de métodos estadísticos.

Si bien inicialmente el término genérico ‘meta-análisis’ se utilizó para designar a toda revisión sistemática, en la práctica son dos métodos distintos. Todo meta-análisis constituye una revisión sistemática de la literatura, pero no toda revisión sistemática se puede considerar un meta-análisis, ya que la primera puede tener un carácter cualitativo. A colación de esto último, cabe decir que aunque las fases del proceso de ambos métodos son las mismas, el meta-análisis aplica técnicas estadísticas para integrar los resultados de las investigaciones empíricas y cuantificar el efecto de la relación bajo estudio. Las fases a seguir en la realización de un meta-análisis las desarrolla Sánchez-Meca (2010):

1. *Formulación del problema.* Al igual que ocurre en una investigación primaria, inicialmente se debe estipular el problema objeto de estudio. Para ello, nos formularemos una serie de preguntas que deben ser respondidas a partir de los resultados obtenidos. Una vez que tenemos claro cuál es el foco de nuestra investigación, se establecerán los objetivos a alcanzar, así como las hipótesis de partida.
2. *Búsqueda de la literatura científica.* Constituye la fase más relevante y laboriosa del proceso. Una vez tengamos clara la pregunta a responder, se hará una búsqueda, lo más exhaustiva posible, de la literatura científica para identificar aquellos estudios de carácter cuantitativo o empíricos cuyo objetivo se corresponda con el nuestro. Para ello, a partir de una serie de palabras clave que utilizaremos como motor de búsqueda, haremos uso de bases de datos electrónicas, así como de meta-busadores para la identificación de estudios. Asimismo, se harán búsquedas manuales en libros y revistas científicas, se rastrearán las referencias bibliográficas de artículos primarios ya seleccionados para identificar otros potenciales estudios más antiguos (*ancestry approach*), y se contactará con investigadores especialistas en la temática bajo estudio con el objetivo de identificar artículos no publicados. Con motivo de ajustar la búsqueda, además del uso de palabras clave, se establecerán unos criterios

de inclusión (e.g., características de la muestra, criterio temporal de publicación en revistas científicas) que deben satisfacer los artículos para ser seleccionados, así como unos criterios de exclusión (e.g., artículos con deficiencias metodológicas o artículos de baja calidad) que les dejarían fuera del meta-análisis. Esta fase debe quedar descrita de forma detallada y rigurosa para que pueda ser sometida a réplica por parte de otros autores.

3. *Codificación de los estudios primarios.* Una vez seleccionados los estudios que formarán parte del meta-análisis, se elaborará un cuadro descriptivo en el que se recogerán las distintas características de las investigaciones primarias (e.g. tamaño de la muestra, existencia o no de grupos control, medida del efecto). La codificación de la información deberá llevarse a cabo por al menos dos jueces independientes para garantizar la calidad de dicha codificación, mediante el cálculo de la fiabilidad inter-jueces (Botella & Gambara, 2002). El número y tipo de características sometidas a codificación dependerá del área de investigación y de la información detallada por las investigaciones previas. La utilidad de la realización de un cuadro base de los estudios es la identificación de variables moderadoras o mediadoras que puedan incidir en los resultados y expliquen la variabilidad observada en los datos.
4. *Cálculo del tamaño del efecto.* Al mismo tiempo que codificamos los estudios, debemos convertir los índices estadísticos utilizados en cada estudio primario (e.g., medias y desviaciones típicas,  $t$  para la diferencia de medias, prueba Chi-cuadrado, proporciones) a una métrica común: el tamaño del efecto. El tamaño del efecto es un índice estadístico que nos informa en qué medida el fenómeno que estamos investigando está presente en la población o, lo que es lo mismo, el grado en que la hipótesis nula es falsa. El tamaño del efecto puede ser expresado en distintos índices, siendo los más comunes la  $d$  de Cohen y la correlación de Pearson ( $r$ ). Las fórmulas para la conversión de los diferentes índices estadísticos a tamaños del efecto varían atendiendo a la relación existente entre ambos. Según Rosenthal (1984), la relación entre prueba de significación y tamaño del efecto responde a la siguiente ecuación general:

$$\text{Prueba de significación} = \text{Tamaño del efecto} \times \text{Tamaño del estudio}$$

Las fórmulas para el cálculo del tamaño del efecto (Cohen, 1988; Rosenthal, 1984) deben seleccionarse meticulosamente, en función las pruebas estadísticas disponibles en los estudios primarios (e.g., medias y desviaciones típicas,  $t$  para una muestra,  $t$  para muestras relacionadas, tabla de contingencia 2X2,  $F$  con un grado de libertad en el numerador).

5. *Análisis estadístico e interpretación.* Finalizado el proceso de codificación y el cálculo del tamaño del efecto para cada estudio, se procederá a elaborar una base de datos para cada investigación primaria. Previamente a la realización del meta-análisis propiamente dicho, algunos autores aconsejan rastrear los datos incorporados en la base de datos en busca de valores outliers o extremos que puedan condicionar la variabilidad observada entre los estudios. Durante el proceso de depuración, debemos ser cautos ya que podemos estar eliminando estudios moderadores con apariencia de valores outliers. Las técnicas de análisis estadístico varían en función del procedimiento meta-analítico. Las modalidades de revisiones meta-analíticas se abordarán con mayor profundidad en un apartado posterior.

Por lo que respecta a la interpretación de los efectos obtenidos, no existen unas reglas absolutas para interpretar la magnitud del fenómeno, ya que depende del área de investigación en la que se encuadre nuestro estudio. Sin embargo, en la comunidad científica se aceptan unos valores establecidos convencionalmente que varían según el índice de tamaño del efecto que hayamos utilizado. Concretamente, ciñéndonos a la clasificación establecida por Cohen (1988), hablaremos de un tamaño del efecto pequeño ( $d = 0.20$ ;  $r = .10$ ), moderado ( $d = 0.50$ ;  $r = .30$ ) o grande ( $d = .80$ ;  $r = .50$ ).

Adicionalmente, en los artículos meta-analíticos aquí presentados, se han calculado otros índices que facilitan al lector la interpretación de los efectos hallados:

- i. Binominal Effect Size Display (BESD). Se trata de un índice que permite conocer la importancia o utilidad práctica del efecto estimado a partir de una diferencia de proporciones. El BESD informa del efecto de una variable predictora sobre la tasa de éxito o mejora en la variable criterio, y se expresa como la diferencia entre el porcentaje de éxito de la tasa base y el porcentaje de éxito tras la aplicación de la variable a medir (Rosenthal, 1984). En nuestro caso concreto, permite calcular la probabilidad de falsos positivos (identificar como víctima a quien no lo ha sido) y falsos negativos (no identificar a víctimas reales).
- ii. Probabilidad de superioridad (PS). También denominado Common Language Effect Size Statistic (CLES) (McGraw & Wong, 1992), hace referencia a la probabilidad de que un individuo extraído al azar de una población obtenga un valor superior que otro individuo, perteneciente a otra población, extraído del mismo modo. En nuestro caso concreto, el cálculo de PS nos permitió conocer en porcentaje de declaraciones sobre hechos auto-experimentados que contenían más criterios de realidad que las declaraciones fabricadas.
- iii. Estadístico U1 (Cohen, 1988). Medida del área, expresada en porcentaje, que no se superpone entre dos distribuciones poblacionales. Es decir, el porcentaje de no-superposición ( $100\% - U1$ ) indica en qué medida un experimento o intervención han tenido un efecto separador entre las dos puntuaciones o poblaciones de interés. Así, cuando  $d = 0.00$ ,  $U1 = 0.00\%$ , las dos poblaciones se superponen al 100% lo que indicaría que ambas poblaciones son idénticas.

Estos dos últimos índices ayudan al lector a comprender mejor la relación entre las distribuciones de las condiciones estudiadas, y se recomienda su inclusión en el informe meta-analítico conjuntamente al del tamaño del efecto para aquellos resultados más relevantes (Fritz, Morris, & Richler; 2012).

6. Publicación del estudio meta-analítico. La culminación de la revisión meta-analítica tiene lugar con su publicación en una revista científica. Puesto que estamos ante una investigación empírica, los apartados que debe contener el artículo son los mismos que los de una investigación primaria: introducción, método, resultados y discusión.

Como ya se explicó anteriormente, la revisión sistemática y el meta-análisis siguen el mismo proceso de elaboración. Sin embargo, a diferencia de lo que ocurre en el meta-análisis, las etapas relativas al cálculo del tamaño del efecto y al análisis estadístico no se llevan a cabo en una revisión sistemática.

### 1.6.1. Tipos de meta-análisis

Los modelos estadísticos para la combinación de resultados que más atención han recibido en la teoría meta-analítica han sido los modelos de efectos fijos y de efectos aleatorios. El primero de ellos asume que no existe heterogeneidad entre los estudios primarios incluidos en la revisión meta-analítica, de modo que si se observa variabilidad, ésta se debe exclusivamente al error de muestreo intra-estudio. Por otro lado, en el modelo de efectos aleatorios se acepta la posibilidad de que los parámetros poblacionales varíen entre estudios, esto es, se asume a priori una heterogeneidad intra e inter-estudios.

Existen diferentes métodos meta-analíticos en función del índice estadístico que se utilice para la acumulación de resultados. Hunter y Schmidt (2015) se centran en aquellos que utilizan correlaciones o tamaños del efecto, dejando de lado las metodologías basadas en los valores del test de significación ( $p$ ). Atendiendo a esta distinción, Hunter y Schmidt (2015) proponen la siguiente clasificación de métodos meta-analíticos: puramente descriptivos, cuyo objetivo es dibujar el panorama sobre una determinada cuestión general; aquellos que sólo tienen en cuenta el error de muestreo (i.e., *bare-bones* meta-análisis); y métodos que tienen en cuenta y corrigen por el error de muestreo y otros artefactos (e.g., meta-análisis psicométrico).

El meta-análisis psicométrico y el *bare-bones* meta-análisis, metodologías elaboradas por Hunter y Schmidt, se enmarcan en los modelos de efectos aleatorios. Estos modelos asumen que los parámetros poblacionales pueden variar entre estudios y que el propósito último es estimar dicha variabilidad. De hecho, el modelo de base es sustractivo, esto es, la varianza poblacional estimada es la varianza que resulta de la eliminación del error de muestreo y otros artefactos (Hunter & Schmidt, 2015). Según estos autores, el procedimiento denominado *bare-bones* es incompleto e insatisfactorio, debido a que solamente corrige el tamaño del efecto estimado por el error de muestreo cuando es habitual que existan otros errores artificiales que alteran el valor de las medidas. Como consecuencia, Hunter y Schmidt proponen los modelos psicométricos como una alternativa más robusta, puesto que permiten conocer si la variabilidad observada entre los estudios se debe a la presencia de artefactos (e.g. error de muestreo, error de medida, restricción en el rango, dicotomización de una variable) o si se trata de variabilidad real. Además, los artefactos crean sesgos a la baja, es decir, infravaloran los resultados del meta-análisis lo que se traduce en una estimación del tamaño del efecto inferior al que es en realidad.

Los métodos de meta-análisis psicométrico tienen dos variantes:

- a) *Meta-analysis of correlations or experimental effect sizes corrected individually for artifacts*: Meta-análisis que corrige individualmente cada  $r$  o  $d$  por los diferentes artefactos. Con frecuencia, no disponemos de información relativa a artefactos en todos los estudios primarios que van a formar parte del meta-análisis. Por ello, los autores plantean una segunda tipología de meta-análisis psicométrico:

- b) *Meta-analysis of correlations or experimental effects sizes using artifact distributions*. Meta-análisis que corrige el tamaño del efecto por la distribución de artefactos. Ante la falta de información requerida para corregir individualmente cada estudio, se opta por corregir a partir de una distribución de los artefactos disponibles. Las revisiones meta-analíticas de las que consta esta tesis, siguen este modelo de procedimiento.

Hunter y Schmidt (2015) identificaron una serie de errores artificiales que alteran el efecto observado en comparación con el efecto real o verdadero:

- 1) Error de muestreo: está relacionado con el tamaño de la muestra, esto es, a menor tamaño muestral, mayor será el error de muestreo. Este artefacto provoca que el tamaño del efecto estimado varíe aleatoriamente del valor de la población.
- 2) Error de medida: puede observarse tanto en la variable dependiente (criterio) como en la variable predictora o independiente: se trata del error de medida aleatorio provocado por la falta de fiabilidad (unreliability) del instrumento de medida. Cuando existe este error el tamaño del efecto observado es sistemáticamente menor que la validez operativa.
- 3) Dicotomización de una variable continua: consiste en transformar los distintos valores de una variable continua a sólo dos, y puede ocurrir tanto en la variable independiente como en la dependiente. La consecuencia sería una atenuación del tamaño del efecto verdadero en comparación con el que obtendríamos con la variable continua.
- 4) Variaciones en el rango: cuanto menos variabilidad exista en la población, es decir, cuanto más homogéneas sean las puntuaciones, mayor será la restricción en el rango y menor el tamaño del efecto observado en relación con el efecto verdadero. Afecta a las variables predictora y criterio.
- 5) Errores de información o transcripción: hace referencia a los fallos cometidos durante el volcado de información en una base de datos, el uso de fórmulas incompletas o incorrectas (errores tipográficos), cuando hacemos una transcripción equivocada o la inexactitud de información al recuperarla de la memoria. Son errores frecuentes y, probablemente, los más difíciles de detectar y corregir posteriormente.
- 6) Varianza debida a factores extraños: se entiende por factores extraños aquellos casos que se desvían significativamente del resto y que pueden hacer variar los resultados. Se les conoce comúnmente como valores outliers y la dificultad estriba en diferenciarlos de valores extremos que son propios de la población de referencia.

Al igual que cualquier otro procedimiento de análisis estadístico, las técnicas meta-analíticas no están exentas de inconvenientes. Primero, existe una alta propensión a cometer errores cuando incorporamos a nuestro meta-análisis investigaciones primarias con sesgos como, por ejemplo, errores cometidos durante el registro de los datos o en el proceso de análisis de los mismos. Segundo, la limitación anterior está íntimamente ligada con la calidad final de los datos. Estudios empíricos con deficiencias metodológicas pueden sesgar los resultados, por ello, ciertos autores son partidarios de eliminar a priori las investigaciones de baja calidad, esto es, tomando la calidad metodológica del estudio como un criterio de

inclusión. Sin embargo, en el otro polo, se apoya la inclusión inicial de todos los estudios empíricos seleccionados, independientemente de la calidad de sus datos y, sólo a posteriori, tomar la calidad de cada estudio como variable moderadora (Hunter & Schmidt, 2015) o bien se someten los resultados a un análisis de sensibilidad. Tercero, resulta imposible abarcar todo el material empírico existente sobre la temática de interés, ya sea porque no se han localizado algunos estudios con la metodología de búsqueda bibliográfica utilizada, o por las dificultades que entraña la localización de investigaciones no publicadas. Esta falta de representatividad de los estudios localizados en relación a la totalidad de los estudios existentes, nos puede llevar a incurrir en el denominado sesgo de publicación (*File-Drawer Problem, FDA*), constituyendo una amenaza a la validez de las conclusiones. Se recomienda, por tanto, el cálculo del *número de seguridad o tolerancia*, es decir, el número de estudios con resultado no significativo que serían necesarios para invertir la conclusión del efecto obtenido. De este modo se obtendrá una aproximación de la amenaza del sesgo de publicación a la validez de las conclusiones alcanzadas. Cuarto, si se toma más de un tamaño del efecto calculado a partir de la misma muestra, no garantizamos la independencia estadística de los datos, lo que conlleva una disminución de la fiabilidad de los resultados. Ante esta situación, se recomienda hallar un promedio de los tamaños del efecto del estudio correspondiente (Sánchez-Meca, 1986). Quinto, la inclusión de estudios muy heterogéneos entre sí, es decir, difícilmente comparables, pone en entredicho la generalización de los resultados.

Por otro lado, las técnicas meta-analíticas poseen una serie de ventajas frente a las revisiones narrativas y a las investigaciones experimentales primarias. En primer lugar, se trata de un procedimiento eficiente en sí mismo, pues permite integrar los resultados de un gran número de estudios que versan sobre una misma temática de forma sistemática y objetiva, permitiendo así su replicación. En definitiva, el meta-análisis goza del mismo rigor científico que un estudio empírico, al exigírsele las mismas normas. En segundo lugar, la precisión (en términos de fiabilidad y validez) de las conclusiones es superior a la de las revisiones cualitativas, dotando al meta-análisis de capacidad para detectar pequeños efectos. En tercer lugar, ostentan una mayor potencia empírica al combinar los resultados de los artículos, aumentando el tamaño muestral y, por ende, la potencia estadística de la prueba aplicada (Sánchez-Meca, 1986). En cuarto lugar, no solamente permite conocer el estado de la cuestión de la problemática estudiada, sino que también detecta áreas de incertidumbre que es necesario abordar en futuras líneas de investigación. Por último, la existencia de resultados contradictorios en la literatura previa pueden ser abordados mediante el estudio de variables moderadoras, las cuales podrían explicar esas discrepancias inter-estudios.



## 2. OBJETIVOS

La elección del procedimiento meta-analítico como el método estadístico más adecuado para la consecución de los objetivos establecidos en esta tesis doctoral radica en las ventajas y facilidades del propio método. A continuación se exponen, artículo por artículo, los objetivos específicos abordados por la revisión meta-analítica:

1. La investigación científica, a nivel cualitativo y cuantitativo, ha arrojado resultados aparentemente contradictorios e interpretaciones dispares sobre la capacidad del CBCA como herramienta válida para discriminar entre memorias de hechos auto-experimentados y memorias de hechos inventados o fabricados en poblaciones de menores. Asimismo, Vrij (2008) dejó patente en su revisión cualitativa de la literatura, la falta de consenso entre la comunidad científica (criterio Daubert número 5), en la aceptación del CBCA como prueba científica para ser admitida en los tribunales de justicia. Es decir, se ha puesto en entredicho que el CBCA cumpla los requerimientos legales de admisibilidad de toda prueba científica en la justicia (Daubert Standards y criterios jurisprudenciales).

Ante este estado de la cuestión, la técnica meta-analítica fue la más pertinente para: a) acercar posiciones entre la comunidad científica sobre la aplicabilidad de la técnica de análisis criterial llegando a conclusiones generalizables a la práctica forense a partir de la interpretación de los datos cuantitativos; b) validar el CBCA como instrumento con capacidad para discriminar entre memorias de hechos auto-experimentados y fabricados (cuantificando el efecto de la relación bajo estudio) y, por consiguiente, afianzar su aplicabilidad en contextos forenses y en diferentes casuísticas (a partir de la generalización de los resultados); c) poner a prueba la científicidad de la técnica a partir del estudio del cumplimiento de los criterios Daubert; d) contribuir al desarrollo de la teoría del testimonio, concretamente al estudio de la credibilidad a partir del análisis del contenido de la declaración, gracias al aporte de la técnica meta-analítica como instrumento de acumulación sistemática de conocimiento científico; d) proponer futuras líneas de investigación que aporten nuevo conocimiento sobre la materia bajo estudio.

2. La utilización de las técnicas de análisis de contenido del testimonio ha sufrido un aumento exponencial en la justicia española, especialmente cuando se trata de evaluar el testimonio de víctimas adultas. La ausencia de validación del CBCA con población adulta, ha generado la reactancia de ciertos sectores del ámbito judicial y forense para su aplicación en dicha población y en contextos distintos para que el que fue creado en origen (i.e., menores víctimas de abuso sexual).

Este panorama puso en duda al CBCA como método categorial válido para ser aplicado en el contexto judicial, desamparando a las víctimas ante la ausencia de una metodología

fiable y válida para la evaluación de la credibilidad del testimonio. Igualmente, genera en los jueces un malestar por la incertidumbre en la valoración de la cientificidad de las pruebas periciales y la validez de la metodología empleada.

De nuevo nos encontramos que, para dar una respuesta eficaz al panorama descrito, la técnica meta-analítica se erige como la más adecuada, permitiendo: a) integrar los resultados de un amplio surtido de investigaciones empíricas con el objetivo de poner a prueba la Hipótesis Undeutsch en población adulta y, por extensión, los criterios de realidad; b) validar la aplicabilidad de la técnica en el sistema de justicia a partir del análisis empírico de los criterios Daubert; c) abordar los resultados contradictorios de los estudios primarios a partir del estudio de variables moderadoras que expliquen la variabilidad inter-estudio observada; d) abordar las casuísticas bajo las cuales los resultados son generalizables y, por tanto, aplicables de forma significativa a otras muestras y a otros actores implicados (i.e., testigos, agresores), arrojando nuevo conocimiento y aportando datos para la construcción de teoría; e) detectar áreas de incertidumbre para la construcción de futuras líneas de investigación; f) consecución de resultados precisos, en términos de fiabilidad y validez.

3. La identificación de la huella psíquica o malestar emocional como consecuencia de la victimización de un hecho delictivo es clave para el cumplimiento del requerimiento legal de la relación causa-efecto. La respuesta psicológica que la literatura ha identificado sistemáticamente con la huella psíquica ha sido el Trastorno por Estrés Postraumático (TEP) (Arce & Fariña, 2009). Sin embargo, la elevada prevalencia de las medidas indirectas del TEP, i.e., la depresión y ansiedad, en menores víctimas de abusos sexuales, bien de forma aislada o comórbida al TEP, se han identificado como las medidas de la huella psicológica consecuencia del hecho traumático, afectando a diversas áreas de la vida de la víctima y provocando un desajuste psicológico en la etapa adulta. Por ello, se ha decidido estudiar la sintomatología ansiosa y/o depresiva, así como el diagnóstico de ansiedad o depresión como medida de la huella psicológica consecuencia del delito.

La metodología que nos permite llegar a conclusiones precisas y generalizables en relación con nuestro objetivo fue, de nuevo, la revisión meta-analítica. A partir de este método se pretendía: a) integrar los resultados de un vasto número de investigaciones empíricas, facilitando así la interpretación de los resultados; b) cuantificar la magnitud de la relación (huella psíquica) entre victimización de ASI/ASA y secuelas internalizantes a partir del cálculo del tamaño del efecto de la relación estudiada; c) arrojar luz sobre la influencia de determinadas variables (p.e. el sexo de la víctima, severidad del abuso) en la magnitud del sufrimiento emocional percibido por las víctimas; d) abordar la cuantificación del daño para aportar conocimiento a operadores políticos y jurídicos, sobre la relevancia de adecuar las penas a la magnitud del daño ocasionado; e) proporcionar una estimación, en términos de probabilidades, de las secuelas clínicas que el o la menor desarrollarán en la adultez como consecuencia de la victimización; f) llegar a conclusiones más precisas y generalizables, gracias a la potencia estadística con la que cuenta el meta-análisis (aumenta significativamente el tamaño muestral); g) finalmente, detectar pequeños efectos que pueden ser altamente relevantes en el campo de la Psicología Forense, concretamente, para el establecimiento de la relación causa-efecto entre el delito y el malestar psicológico manifestado, especialmente cuando se trata de un colectivo vulnerable.

### 3. MÉTODO

El método que se ha seguido para la consecución de los objetivos anteriormente expuestos, fue la revisión meta-analítica, concretamente, se ha utilizado el procedimiento meta-analítico elaborado por Hunter y Schmidt (2015).

Los pasos a seguir para la realización de los meta-análisis de experimentos, esto es, aquellos en los que el tamaño del efecto consistió en la  $d$  de Cohen (1988) fueron los que siguen. Primero, se llevó a cabo una búsqueda exhaustiva y multimétodo, con el objetivo de identificar estudios empíricos que pusieran a prueba la capacidad discriminativa de los criterios de realidad del CBCA (tanto en muestras de menores como en población adulta). Para ello, hemos recurrido a bases de datos internacionales de referencia científica utilizando una serie de palabras clave como motor de búsqueda en distintos idiomas (i.e., inglés, castellano, italiano, francés, alemán). Asimismo, se ha completado la búsqueda en metabuscadores como Google Scholar; en redes sociales académicas para el intercambio de conocimiento dentro de la comunidad científica (i.e., Researchgate, Academia.edu); se ha seguido el procedimiento denominado *ancestry approach*, esto es, el cotejo de las listas de referencia de los trabajos seleccionados; y finalmente, se ha contactado con investigadores para la solicitud de estudios no publicados, mencionados en otras investigaciones.

Una vez identificados los potenciales estudios, se aplicaron los criterios de inclusión y de exclusión, de acuerdo con los objetivos de cada uno de los meta-análisis, de manera que se pudiera obtener estudios homogéneos en sus características. Una vez seleccionados los estudios primarios que formarían parte del meta-análisis, se procedió a su clasificación en una rejilla de codificación, en función de aquellas variables que pudieran ejercer un rol de moderador.

Los tamaños del efecto fueron tomados directamente de los estudios primarios cuando informaban de ellos, en caso contrario, se calcularon a partir de las medias y desviaciones estándar/errores estándar: tamaño del efecto  $d$  de Cohen cuando  $N_1 = N_2$  y delta de Glass ( $\Delta$ ) cuando  $N_1 \neq N_2$ . Cuando los resultados estaban expresados en  $t$  de student o eta cuadrado fueron transformados a  $d$  de Cohen utilizando las fórmulas de conversión de Rosenthal (1994) y Cohen (1988). En caso de que estuvieran expresados en valores  $F$  de Fisher sin la probabilidad exacta, el valor de  $F$  fue transformado a  $t$  y posteriormente a  $d$ . Cuando los resultados estaban en proporciones se obtuvo el tamaño del efecto  $\delta$  (Hedges & Olkin, 1985), que se corresponde con  $d$ , en tanto que cuando se facilitaban en tablas de contingencia 2X2 se obtuvo phi y, posteriormente,  $d$ .

Las diferencias entre tamaños del efecto se estimaron con el estadístico  $q$  de Cohen a partir de la transformación de los tamaños del efecto  $d$  en correlaciones ( $r$ ). Adicionalmente, excediendo la finalidad propia de un procedimiento meta-analítico, y con el objetivo de

conocer la utilidad práctica de los resultados en el campo forense, se emplearon los estadísticos U1, Binomial Effect Size Display (BESD) y Probabilidad de Superioridad (PS).

Finalmente, para apoyar la robustez de los efectos encontrados, se procedió al cálculo del sesgo de publicación o *File Drawer Analysis* (FDA; McNatt, 2000), que permite conocer el número de estudios no detectados con efecto nulo que serían necesarios para reducir los efectos obtenidos en el meta-análisis a uno trivial (.05), es decir, próximo al cero. Si el valor obtenido es elevado (la interpretación de dicho valor dependerá del área de estudio), podremos concluir que el sesgo de publicación no es una amenaza contra la validez de los resultados.

Como quiera que el procedimiento meta-analítico de Hunter y Schmidt permite conocer la influencia en los resultados de errores artificiales y, consecuentemente, corregir el tamaño del efecto observado por dichos errores, se ponderó cada  $d$  por el tamaño de la muestra y se corrigió por la fiabilidad del criterio. En los meta-análisis en los que se puso a prueba la validez de los criterios de realidad del CBCA, se computó la fiabilidad del criterio a partir de las fiabilidades (no así los índices de acuerdo) inter-jueces (i.e., codificación de los criterios de realidad como presentes o ausentes en las declaraciones analizadas) aportadas por cada estudio independiente. Se estimó una fiabilidad promedio para los criterios de realidad, debido a la falta de información de la fiabilidad en la codificación en los estudios primarios. Por otro lado, se estimó una fiabilidad promedio para la puntuación total del CBCA, ya que no es la misma que la de los criterios individuales (Horowitz et al., 1997), utilizando para ello la fórmula de Spearman-Brown.

$$r_{xx} = \frac{n * \bar{r}_{cxx}}{1 + ((n-1) * \bar{r}_{cxx})}$$

Previamente al cálculo de los correspondientes meta-análisis, se examinaron los tamaños del efecto con la finalidad de identificar posibles valores outliers o extremos. En el caso de detectar algún outlier, se analizaban las características del estudio para asegurarnos de que no era un moderador. En caso de tratarse de un valor outlier, se procedía a su eliminación, siempre y cuando no supusiera eliminar más de un 10% de los tamaños del efecto de cada meta-análisis (Tukey, 1960).

Además de lo anterior, se computó un intervalo de confianza para cada meta-análisis con el fin de conocer si el efecto estimado era estadísticamente significativo, y un intervalo de credibilidad para conocer si el efecto estimado podría generalizarse (escasa variabilidad en los datos) a otras poblaciones. Finalmente, la interpretación del porcentaje de varianza explicada por errores artificiales (%Var) está íntimamente ligada al intervalo de credibilidad, ya que ambos informan sobre la heterogeneidad entre estudios y de la necesidad de analizar variables moderadoras que puedan estar incidiendo en los resultados e impidiendo su generalización a otras muestras y contextos.

Como resultado, se mostró para cada meta-análisis una tabla informativa que contenía los siguientes datos (presentación de los datos en la tabla de izquierda a derecha): el número de tamaños del efecto ( $k$ ), el tamaño de la muestra ( $n$ ), desviación estándar observada de  $d$  corregida por todos los artefactos ( $SD_{pre}$ ); desviación estándar observada de  $d$  después de eliminar la varianza debida a todos los artefactos ( $SD_{res}$ ); tamaño del efecto corregido por falta

de fiabilidad en el criterio ( $\delta$ ); desviación estándar de  $\delta$  ( $SD_\delta$ ); varianza explicada por todos los errores artificiales (%Var); intervalo de confianza para  $d$  al 95% (95%  $CI_d$ ); intervalo de credibilidad para  $\delta$  al 80% (80%  $CV_\delta$ ) o al 90% (90%  $CI_\delta$ ).

Con respecto al procedimiento meta-analítico de correlaciones, donde el tamaño del efecto estimado consistió en la correlación de Pearson ( $r$ ), los pasos a seguir fueron los mismos que en el meta-análisis de experimentos: búsqueda de literatura científica en la base de datos de excelencia científica (Web of Science) utilizando una metodología multimétodo; establecimiento de criterios de inclusión y de exclusión; identificación y codificación de los estudios primarios que formaron parte, finalmente, de los meta-análisis; cálculo de los tamaños del efecto (cuando no eran proporcionados directamente por los estudios primarios); búsqueda de valores outliers; y cálculo del propio meta-análisis. En esta parte del proceso, la única diferencia sustancial respecto del meta-análisis de experimentos, fue la corrección del tamaño del efecto observado no solo por el error de muestreo y el error de medida en la variable dependiente (falta de fiabilidad en el criterio), sino también el error de medida en la variable independiente (falta de fiabilidad en el predictor, esto es, la medida de la victimización).

La tabla de resultados del meta-análisis de correlaciones, informaba de los siguientes indicadores: número de tamaños del efecto independientes ( $k$ ); tamaño muestral del grupo experimental ( $N_E$ ); tamaño muestral del grupo control ( $N_C$ ); tamaño muestral total ( $N_T$ ); correlación observada (validez observada) ponderada por el tamaño de la muestra ( $r_w$ ); desviación estándar de la correlación observada ( $SD_r$ ); correlación verdadera (validez operativa) corregida por falta de fiabilidad en el predictor y en el criterio, ( $\rho$ ); desviación estándar de la correlación verdadera ( $SD_\rho$ ); porcentaje de varianza explicada por los errores artificiales (%VE); intervalo de confianza al 95% (95%  $CI_r$ ); intervalo de credibilidad al 90% (90%  $CI_\rho$ ).



## 4. RESULTADOS

De los meta-análisis efectuados se desprenden los siguientes resultados:

1. Se ha obtenido un tamaño del efecto corregido por falta de fiabilidad en el criterio positivo y grande ( $\delta = 0.79$ ), significativo (el intervalo de confianza no contiene el valor cero) y generalizable, en la puntuación global del CBCA, en muestras de menores. En muestras de adultos, el tamaño del efecto para la puntuación total del instrumento ha sido de magnitud moderada ( $\delta = 0.56$ ), significativo, pero no generalizable a otras poblaciones (el intervalo de credibilidad contiene el valor cero). El tamaño del efecto hallado para el total del instrumento fue significativamente superior (los intervalos de confianza para los valores  $\delta$  no se superponen; ver Tabla 2) en la muestra de menores que en la de adultos.
2. En relación con los criterios individuales de realidad del CBCA, todos los criterios han sido positivos (el rango de los tamaños del efecto oscila entre  $\delta = 0.17$  para el criterio 'admisión de falta de memoria', y  $\delta = 1.40$  en el caso del criterio 'detalles característicos de la agresión'), significativos, esto es, con capacidad para discriminar entre memorias de menores basadas en hechos auto-experimentados y memorias de hechos inventados, sin embargo, no son generalizables a otras muestras. Por otro lado, los resultados hallados en muestras de adultos informan de tamaños del efecto positivos (salvo el criterio 'perdón al autor del delito',  $\delta = -0.02$ ) y significativos, es decir, los criterios de realidad del CBCA poseen capacidad discriminativa (con excepción de los criterios 'auto-desaprobación' y 'perdón al autor del delito' que no fueron productivos, el intervalo de confianza contenía el valor cero) pero no generalizables. Los tamaños del efecto corregidos por falta de fiabilidad en el criterio oscilaron entre  $\delta = -0.02$  para el criterio 'perdón al autor del delito' y  $\delta = 0.71$  para 'cantidad de detalles'. En general, la mayoría de los criterios de realidad discriminan por igual tanto en muestras de menores como en adultos, salvo los criterios 'engranaje contextual', 'detalles superfluos', 'estado mental subjetivo' y 'detalles característicos de la ofensa', que discriminan significativamente mejor las declaraciones basadas en hechos vividos frente a los hechos inventados en población de menores (ver Tabla 2).

Tabla 2. Estudio de la eficacia discriminativa diferencial de los criterios de realidad y del total, en muestras de menores y adultos

Criterios	$k_1$	$k_2$	$N_1$	$N_2$	$\delta_1$	$\delta_2$	$DE_1$	$DE_2$	$IC_{\delta_1} 95\%$	$IC_{\delta_2} 95\%$
Estructura lógica	16	30	1381	2265	0.52	0.62	0.272	0.849	[0.41, 0.63]	[0.53, 0.70]
Producción inestructurada	15	27	1217	1987	0.53	0.69	0.588	1.155	[0.42, 0.64]	[0.60, 0.78]
Cantidad de detalles	17	35	1477	2714	0.87	0.71	0.594	1.028	[0.76, 0.97]	[0.63, 0.78]
Engranaje contextual	15	29	1341	2137	0.78	0.24	0.571	0.741	[0.67, 0.89]	[0.15, 0.32]
Descripción de interacciones	16	29	1407	2243	0.50	0.36	0.434	0.379	[0.39, 0.60]	[0.28, 0.44]
Reproducción de conversaciones	16	34	1407	2528	0.59	0.44	0.415	0.607	[0.48, 0.69]	[0.36, 0.52]
Complicaciones inesperadas	11	29	1111	1956	0.33	0.32	0.000	0.370	[0.21, 0.45]	[0.23, 0.41]
Detalles inusuales	16	35	1437	2441	0.31	0.41	0.335	0.786	[0.21, 0.41]	[0.33, 0.49]
Detalles superfluos	13	27	1199	1863	0.47	0.18	0.274	0.667	[0.35, 0.58]	[0.09, 0.27]
Detalles incomprensidos	13	5	1062	376	0.35	0.28	0.385	0.000	[0.23, 0.47]	[0.08, 0.48]
Asociaciones externas	10	22	916	1612	0.32	0.34	0.328	0.538	[0.19, 0.45]	[0.24, 0.44]
Estado mental subjetivo	15	28	1194	2170	0.52	0.23	0.419	0.554	[0.40, 0.63]	[0.14, 0.31]
Estado mental autor del delito	10	31	1052	2232	0.21	0.11	0.222	0.747	[0.09, 0.33]	[0.03, 0.19]
Correcciones espontáneas	15	29	1367	1842	0.23	0.20	0.383	0.601	[0.12, 0.33]	[0.07, 0.33]
Admisión de falta de memoria	13	34	1076	2305	0.17	0.32	0.318	0.377	[0.05, 0.29]	[0.24, 0.40]
Dudas propio testimonio	10	26	809	1755	0.22	0.26	0.238	0.492	[0.08, 0.36]	[0.16, 0.35]
Autodesaprobación	5	13	447	948	0.18	0.05	0.480	0.519	[-0.01, 0.37]	[-0.08, 0.18]
Perdón al autor del delito	6	8	517	680	0.25	-0.02	0.343	0.228	[0.08, 0.42]	[-0.17, 0.13]
Detalles característicos agresión	5	5	318	562	1.40	0.36	0.807	0.000	[1.15, 1.64]	[0.19, 0.53]
Total	18	31	1122	2124	0.79	0.56	0.275	0.638	[0.67, 0.91]	[0.47, 0.65]

Note.  $k_1$  = número de estudios con muestras de menores;  $k_2$  = número de estudios con muestras de adultos;  $N_1$  = tamaño de la muestra en menores;  $N_2$  = tamaño de la muestra en adultos;  $\delta_1$  = tamaño del efecto en menores corregido por falta de fiabilidad en el criterio;  $\delta_2$  = tamaño del efecto en adultos corregido por falta de fiabilidad en el criterio;  $DE_1$  = desviación estándar de  $\delta_1$ ;  $DE_2$  = desviación estándar de  $\delta_2$ ;  $IC_{\delta_1} 95\%$  = Intervalo de confianza al 95% para  $\delta_1$ ;  $IC_{\delta_2} 95\%$  = Intervalo de confianza al 95% para  $\delta_2$



3. Los criterios adicionales ‘estilo de la declaración’ y ‘dar muestras de inseguridad’, puestos a prueba en el meta-análisis con población adulta, mostraron un tamaño del efecto corregido positivo ( $\delta = 0.48$  y  $\delta = 0.78$ , respectivamente) significativo y generalizable. No obstante, el criterio ‘dar explicaciones de la falta de memoria’ no resultó productivo (el intervalo de confianza contiene el valor cero). Finalmente, los restantes criterios adicionales analizados arrojaron un tamaño del efecto negativo, aunque sólo fue significativo para el criterio ‘repeticiones’ ( $\delta = -0.54$ ), apuntando que no se trata de un criterio de realidad, sino un indicador de engaño.
4. En tanto que la literatura analizada advertía de la existencia de características en los estudios primarios que podrían afectar a los resultados hallados, se llevaron a cabo nuevos meta-análisis con el objetivo de conocer cómo afectaban al efecto estimado. La primera de ellas y, desde el punto de vista de la práctica forense, la más relevante, se refiere al paradigma de investigación (investigación de campo vs. investigación experimental de laboratorio).
  - 4.1. El tamaño del efecto hallado en los estudios de campo (tomando la puntuación total del CBCA) con muestras de menores fue positivo, significativo, generalizable, y de una magnitud más que grande  $\delta = 2.71$ ;  $p < .01$ . Igualmente, con la población de adultos se obtuvo un tamaño del efecto positivo y moderado ( $\delta = 0.69$ ), significativo y generalizable, con el promedio de criterios significativos (los criterios realmente discriminativos entre memorias de hechos vividos y fabricados son los que formarán parte del sistema categorial) como variable dependiente. En resumen, en estudios de campo, el total del instrumento y los criterios de realidad significativos, discriminan significativamente entre memorias de hechos auto-experimentados y memorias fabricadas, tanto en muestras de menores como en adultos, respectivamente.
  - 4.2. Los resultados encontrados para el meta-análisis de estudios experimentales revela un tamaño del efecto positivo y significativo en ambos tipos de muestras (menores,  $\delta = 0.56$ ; y adultos,  $\delta = 0.32$ ), no resultando generalizable a otras poblaciones cuando se aplica el CBCA a adultos. Tomando en consideración las directrices de Steller (1989) para incrementar la validez ecológica de los paradigmas experimentales, se seleccionaron aquellos estudios primarios que cumplieran las condiciones de ‘high fidelity’ (involucración personal en los hechos, con un carácter emocional negativo y que implicaran una pérdida de control respecto del evento) y se llevó a cabo un nuevo meta-análisis. Los resultados hallados fueron los mismos que tomando todos los estudios experimentales ( $\delta = 0.58$  vs.  $\delta = 0.56$ , respectivamente) en la población de menores.
  - 4.3. Igualmente, se ha investigado el efecto del contexto en la capacidad discriminativa de los criterios de realidad en población de adultos. Para ello, se realizó un nuevo meta-análisis tomando únicamente los estudios de campo en casos de violencia sexual y de género (delitos cometidos en la esfera privada). Los resultados revelaron (con el promedio de criterios significativos) un tamaño del efecto elevado ( $\delta = 0.96$ ), positivo,

significativo y generalizable; siendo el efecto más pequeño esperable de  $\delta = 0.64$ . La diferencia de tamaños del efecto, entre la condición de estudios de campo y de abuso sexual y violencia de género (con los criterios significativos) advierten de una mayor capacidad discriminativa de los criterios de realidad bajo la condición de contexto,  $q_c = .168$ ,  $p < .05$  (0.69 vs. 0.96).

5. El análisis del impacto de otras variables en los resultados nos llevó a proceder con el cálculo de distintos meta-análisis (se tomó el promedio de criterios como variable dependiente) seleccionando como moderadores el criterio de publicación Daubert Estándar en revistas con proceso de revisión por pares (DSCP), la versión del sistema categorial (todos los criterios de realidad vs versión de 14 criterios, estatus del declarante, y tipo de evento [eventos auto-experimentados y eventos observados en vídeo]). En todos ellos se obtuvo un tamaño del efecto positivo y significativo, pero no generalizable (es necesario continuar con la búsqueda de moderadores). Los efectos hallados fueron todos pequeños ( $0.20 > \delta < 0.50$ ), a excepción del meta-análisis para los eventos observados en video (testigo) cuya magnitud fue moderada ( $\delta = 0.51$ ). En los anexos pueden consultarse los resultados de los meta-análisis (para cada criterio individual del CBCA, criterios adicionales, total del instrumento, y promedio) de cada variable moderadora en muestras de adultos (tablas 3-11).
6. El estudio del testimonio de agresores valida de nuevo la hipótesis Undeutsch: todos los criterios de realidad fueron positivos. No obstante, los criterios ‘detalles inusuales’, ‘detalles incomprendidos relatados con precisión’, ‘asociaciones externas relacionadas’, ‘estado mental subjetivo’, ‘correcciones espontáneas’, ‘dudas sobre el propio testimonio’, ‘auto-desaprobación’ y ‘perdón al autor del delito’ no fueron productivos y, por tanto, no pueden aplicarse a muestras de agresores.
7. La victimización de abuso sexual reveló un efecto global moderado ( $\rho = .34$ ), en el padecimiento de secuelas psicológicas (depresión y ansiedad), significativo y generalizable. Concretamente, se obtuvo un tamaño del efecto pequeño y otro moderado en padecimiento de sintomatología depresiva ( $\rho = .28$ ) y ansiosa ( $\rho = .31$ ), respectivamente.
8. Los análisis realizados atendiendo al género de la víctima, revelaron un efecto pequeño, significativo y generalizable, para las mujeres, en la manifestación de depresión y ansiedad ( $\rho = .22$ ). Sin embargo, para los varones se obtuvo un tamaño del efecto pequeño, positivo, significativo y generalizable en la medida de la depresión ( $\rho = .13$ ), pero no generalizable a otras muestras en la medida de ansiedad ( $\rho = .15$ ) (el intervalo de credibilidad contiene el cero).
9. Se llevaron a cabo, además, distintos meta-análisis tomando como variable moderadora el tipo de medida del daño psicológico, esto es, diagnóstico o sintomatología. Los resultados hallados informan de un tamaño del efecto moderado, positivo, significativo y generalizable en el diagnóstico de depresión ( $\rho = .31$ ) y ansiedad ( $\rho = .35$ ). Asimismo, la victimización de ASI/ASA también acarrea un desarrollo significativo, positivo, generalizable, y de un tamaño del efecto pequeño, de sintomatología ( $\rho = .21$ ) ansiosa y depresiva.

10. Tomando en consideración la epidemiología de los trastornos depresivos y los trastornos de ansiedad para cada género (DSM-V), se identificó como moderador la interacción entre el tipo de medida del daño y el género de la víctima.
  - 10.1. Los resultados mostraron un efecto positivo, significativo y generalizable para el diagnóstico de depresión ( $\rho = .42$ ) y ansiedad ( $\rho = .24$ ) en mujeres. El efecto hallado para los varones en ambos diagnósticos también resultó positivo y significativo (aunque de magnitud pequeña), pero no generalizable. Comparativamente, las mujeres víctimas de ASI/ASA son diagnosticadas, de forma significativa, de un trastorno depresivo,  $q_s = 0.388$ ,  $p < .01$ , así como de un trastorno de ansiedad  $q_s = 0.104$ ,  $p < .05$ , en mayor medida que los varones.
  - 10.2. Los efectos hallados para la medida de sintomatología han sido positivos, de un tamaño del efecto pequeño, significativos y generalizables en todos los casos. Concretamente, varones y mujeres presentan efectos similares en el padecimiento de sintomatología depresiva ( $\rho = .22$  y  $\rho = .20$ , respectivamente). No obstante, los varones refieren significativamente más sintomatología ansiosa que las mujeres,  $q_s = 0.095$ ,  $p < .05$ .
11. Los meta-análisis realizados tomando el tipo de abuso o severidad (sin contacto, con contacto y con penetración) como variable moderadora, resultaron en un tamaño del efecto significativo, positivo, de un tamaño pequeño y generalizable, en depresión y ansiedad. La comparación entre los tamaños del efecto mostró que el daño en el abuso con penetración ( $\rho = .19$  en depresión,  $\rho = 0.15$  en ansiedad), era significativamente mayor que en la condición de abuso sin contacto ( $\rho = .12$ ,  $\rho = .08$ ), tanto para depresión,  $q_s = 0.093$ ,  $p < .05$ ; como para ansiedad,  $q_s = 0.092$ ,  $p < .05$



## 5. DISCUSIÓN

De los resultados hallados en los estudios meta-analíticos, se han extraído las siguientes conclusiones.

### **En memorias de menores:**

1. Se validó la Hipótesis Undeutsch en memorias de menores a partir del estudio de los criterios de realidad del CBCA. Esto es, el CBCA como sistema categorial para el análisis de contenido de declaraciones de menores discriminó significativamente entre memorias de hechos vividos y fabricados. Esto se traduce en un mayor número de criterios de realidad en declaraciones basadas en hechos reales frente a las declaraciones inventadas o fabricadas. Por tanto, los resultados apoyan la Hipótesis Undeutsch para niños de todas las edades y contextos (i.e., para contextos ajenos al abuso sexual).
2. La puntuación total en el CBCA discriminó entre declaraciones reales y fabricadas, independientemente del paradigma de investigación (estudios experimentales vs. campo). Sin embargo, la eficacia del sistema resultó significativamente superior en los estudios de campo que en los experimentales. En consecuencia, los resultados de estudios experimentales no son directamente generalizables para la práctica salvo que estén apoyados por estudios de campo.
3. Todos los criterios de realidad del CBCA discriminaron significativamente entre memorias de hechos vividos y fabricados. Esto es, los resultados validaron todos los criterios de realidad del CBCA.
4. Los resultados revelaron tamaños del efecto más elevados para los criterios de realidad del componente cognitivo (criterios 1 al 13 del CBCA) que para el componente motivacional (criterios 14 al 18). Además, los criterios motivacionales no fueron generalizables.
5. La primera gran categoría “características generales” i.e., los criterios ‘estructura lógica’, ‘producción inestructurada’, ‘cantidad de detalles’, han obtenido los tamaños del efecto más grandes. El tamaño del efecto más elevado fue para el criterio ‘detalles característicos de la agresión’. Una explicación radicaría en la dificultad para inventar estos detalles.
6. En cuanto a la práctica forense, la puntuación total en el CBCA clasifica correctamente al 68.5% de las memorias de hechos auto-experimentados (verdaderos positivos), mientras que falla en la detección del 31.5% de dichas memorias (falsos negativos). Respecto a los verdaderos negativos no pueden generalizarse a la práctica puesto que los criterios de realidad no clasifican

memorias falsas/fabricadas. En relación con los falsos positivos (declaraciones falsas identificadas como verdaderas), que en el campo forense debe ser el 0% puesto que transgrede el principio de presunción de inocencia (i.e., el Tribunal Constitucional y el Tribunal Supremo españoles han manifestado que ninguna persona inocente debe estar en prisión, mientras que algunas personas que son culpables puede estar fuera de prisión), la probabilidad potencial de falsos positivos es de .288 (1-.712, es decir, uno menos la probabilidad de obtener más criterios de realidad del CBCA en testimonios de menores relativos a hechos vividos).

7. El análisis de la variable moderadora paradigma de investigación mostró, con la puntuación total del CBCA, un tamaño del efecto y una correcta clasificación de los verdaderos positivos superior en los estudios de campo (90.2%) que en los experimentales (63.5%). Por consiguiente, los resultados derivados de los estudios experimentales no pueden generalizarse directamente a la práctica, pues solamente gozan de validez aparente (Fariña, Arce, & Real, 1994; Konecni & Ebbesen, 1992). Esto corrige también la probabilidad de falsos positivos pasando a ser .029. No obstante, se desconoce el criterio de decisión para alcanzar este resultado, lo que significa que este procedimiento no es una prueba válida cuya decisión se basa en una impresión global o juicio clínico (Arce & Fariña, 2013; Köhnken, 2004).

#### **En memorias de adultos:**

8. Adicionalmente, la Hipótesis Undeutsch fue validada en adultos (puesto que la Hipótesis Undeutsch se sustenta en los contenidos de memoria, ya se había anticipado teóricamente que sería igualmente aplicable tanto en memorias de adultos como en otros contextos ajenos al abuso sexual) (Berliner & Conte, 1993). Asimismo, se validó la Hipótesis Undeutsch no solo para las memorias de víctimas, testigos y agresores, sino también para estudios experimentales (alta validez interna) y de campo (alta validez externa). Sin embargo, los criterios de realidad del CBCA clasificaron significativamente mejor las memorias vividas en estudios de campo que en estudios con diseños de simulación, no siendo éstos generalizables a la práctica (prueba suficiente) por sí mismos.
9. No todos los criterios de realidad fueron válidos (i.e., ‘auto-desaprobación’, ‘perdón al autor del delito’), ni generalizables (i.e., todos los criterios con excepción de ‘detalles incomprendidos relatados con precisión’ y ‘detalles característicos de la agresión’ que estaban afectados por un error de muestreo de segundo orden por lo que sus resultados no fueron válidos para esta estimación). No obstante, el criterio ‘perdón al autor del delito’ fue contrario a la Hipótesis Undeutsch (el tamaño del efecto medio fue negativo).
10. Aunque la Hipótesis Undeutsch y los criterios de realidad son también válidos para memorias de agresores, los resultados no pueden ser indubitablemente extrapolados a la práctica puesto que los datos proceden exclusivamente de estudios experimentales. Además de esto, en la evaluación forense de las memorias reales de agresores no existe la *ground truth* ya que los agresores pueden, por ley, mentir, no declarar, o simplemente negar los hechos alternativamente a contar la verdad. Bajo estas consideraciones, los criterios

‘estructura lógica’, ‘producción inestructurada’, ‘cantidad de detalles’, ‘engranaje contextual’, ‘descripción de interacciones’, ‘reproducción de conversaciones’, ‘complicaciones inesperadas’, ‘detalles superfluos’, ‘estado mental del autor del delito’, ‘admisión de falta de memoria’, ‘detalles característicos de la agresión’ discriminaron significativamente entre memorias reales y fabricadas en agresores, siendo los dos últimos criterios conjuntamente con la puntuación total del CBCA, generalizables a cualquier condición.

11. Los resultados apoyan los criterios de realidad adicionales a los criterios de realidad que componen CBCA (i.e., ‘estilo de la declaración’, ‘dar muestras de inseguridad’, y ‘dar explicaciones de la falta de memoria’). En conclusión, los criterios de realidad del CBCA pueden complementarse con criterios de realidad adicionales.
12. Algunos de los criterios adiciones que fueron estudiados no son criterios de realidad, puesto que estaban inversamente relacionados con memorias vividas. Concretamente, los criterios ‘repeticiones’ y ‘clichés’ mostraron una relación negativa con las memorias vividas. Por tanto, constituyen atributos de memoria, esto es, características de memoria no directamente relacionadas con la realidad del evento, que deben ser estudiados con el objetivo de crear un sistema categorial que combine criterios de realidad y atributos de memoria. De facto, Arce y Fariña (2009; Vilariño, 2010; Vilariño, Novo, & Seijo, 2011) crearon y validaron un sistema categorial para discriminar memorias vividas y fabricadas en víctimas de violencia de género.
13. Los estudios de campo mostraron que no todos los criterios de realidad son válidos para discriminar entre memorias reales y fabricadas. Estos resultados sugieren que el sistema categorial debe reducirse a los criterios de realidad que discriminan significativamente, puesto que los no significativos solamente introducen ruido en la evaluación.
14. En consecuencia, la variable moderadora contexto de evaluación reveló que los criterios de realidad discriminan mejor en los casos reales de agresiones sexual y de violencia de género, clasificando correctamente al 71.5% de los verdaderos positivos y fallando en el 28.5% de los falsos positivos.
15. La eficacia de la puntuación total del CBCA en la discriminación entre memorias vividas y fabricadas fue significativamente superior en la muestra de menores que en la de adultos.
16. Comparativamente, los criterios ‘engranaje contextual’, ‘detalles superfluos’, ‘estado mental subjetivo’ y ‘detalles característicos de la agresión’ mejoran los ratios de clasificación en las muestras de menores. Los restantes criterios tienen la misma eficacia en la clasificación en ambas muestras.

### **Sintomatología internalizante en casos de abuso sexual en la infancia y adolescencia**

17. La victimización de abuso sexual infantil o adolescente (CSA/ASA) conlleva una probabilidad aproximada de .70 de padecer daño, .66 de depresión y .68 de ansiedad. La cuantificación estimada del daño causado fue de un 30% en la secuela en general, depresión y ansiedad.

18. En relación con el diagnóstico de trastornos internalizantes, las mujeres víctimas tienen un 9% más de probabilidades que los hombres de desarrollar depresión, pero no ansiedad. Esto deriva en un elevado daño en depresión (41 vs. 10% en mujeres y hombres, respectivamente) y ansiedad (24 vs. 14%) en mujeres en comparación con los hombres. Este hallazgo es consistente con los resultados de estudios epidemiológicos y de prevalencia informada por organismos internacionales de referencia en la clasificación de enfermedades (APA, 2013; WHO, 2000), señalando la vulnerabilidad de las mujeres en el desarrollo de secuelas internalizantes.
19. La probabilidad de desarrollar un daño crónico (distimia) fue significativamente superior a la probabilidad de un daño más grave o severo (trastorno de depresión mayor).
20. La cuantificación del daño para un trastorno depresivo persistente (distimia) y un trastorno de depresión mayor fue de un 46 y 31%, respectivamente, en las víctimas de ASI/ASA.
21. La victimización de ASI/ASA incrementa en un 43% la probabilidad de desarrollar un trastorno de ansiedad (trastorno de ansiedad generalizada, fobia específica, fobia social y trastorno de pánico).
22. En resumen, la victimización de ASI/ASA predispone a la aparición de problemas de ajuste psicológico en la adultez (Chiesa, Larsen-Paya, Martino, & Trinchieri, 2016).
23. El diagnóstico clínico es significativamente más adecuado (sensible) que la sintomatología clínica para la medida de victimización de ASI/ASA. En consecuencia, el diagnóstico clínico debería ser tomado como medida del daño consecuencia de la victimización de ASI/ASA, en lugar de la sintomatología.
24. La victimización de un delito sexual grave (i.e., con penetración), en contraste con la agresión sin contacto, conlleva un mayor daño en términos de depresión y ansiedad. Esto apoya la adjudicación de penas más altas en los Códigos Criminales cuando el delito de carácter sexual implica penetración, puesto que causa mayor daño psicológico que aquellos delitos contra la libertad e indemnidad sexuales donde la agresión no media contacto.

### **Implicaciones de los resultados para la práctica forense**

La admisibilidad de pruebas científicas en el contexto judicial descansa en el cumplimiento de una serie de criterios establecidos por los tribunales y la jurisprudencia. A este respecto, se presentan a continuación las respuestas a dichos estándares sobre la admisibilidad de los criterios de realidad del CBCA a partir de los resultados alcanzados en los meta-análisis:

1. ¿Puede probarse la hipótesis científica? La hipótesis que subyace al sistema categorial CBCA, la denominada Hipótesis Undeutsch, ha sido sometida a prueba en numerosos estudios (todos los estudios incluidos en el primer, segundo y



cuarto meta-análisis de esta tesis doctoral, sometieron a prueba la hipótesis). Así, la hipótesis científica subyacente puede y ha sido sometida a contraste.

2. ¿Ha sido probada la hipótesis? La hipótesis evaluada a través del CBCA, ha sido probada y firmemente confirmada (con base en los resultados obtenidos en nuestros meta-análisis).
3. ¿Se conoce la tasa de error? Originalmente, no se proporcionaba ni se consideraba una tasa de error (Köhnken, 2004; Steller, 1989; Steller & Köhnken, 1989; Undeutsch, 1989) ya que la técnica era semi-objetiva o basada en el juicio clínico de los expertos. Vrij (2005, 2008) estimó (estudio cualitativo) una tasa de error próxima al 30% para las declaraciones verdaderas y falsas. Sin embargo, los resultados para las memorias falsas no son válidos puesto que la hipótesis y los criterios no clasifican dichas memorias. Nuestros resultados mostraron diferentes tasas de error en la clasificación de memorias reales (falsos positivos), siendo los más significativos, de un 9.8% en casos reales de abuso sexual infantil y de 28.5% en casos reales de violencia de género y delitos de naturaleza sexual en muestras de adultos.
4. ¿Ha sido sometida la hipótesis/técnica a un proceso de revisión por pares y publicación? Sí, muchos estudios que contrastan la hipótesis y la técnica han sido publicados. Muchos de ellos sometidos a un proceso de revisión por pares.
5. ¿Se basa la hipótesis y/o técnica en una teoría generalmente aceptada como válida por la comunidad científica? Aunque la comunidad científica no ha sido consultada como tal, la evidencia científica que procede de investigadores adecuados proporciona soporte (significativo y generalizable en los estudios de campo; ver los resultados de los meta-análisis) a la hipótesis y a la técnica. Puesto que no se trata de un soporte mínimo, en cuyo caso los resultados deben ser considerados con escepticismo (Daubert v. Merrell Dow Pharmaceuticals, Inc., 1993), la conclusión razonable es que la hipótesis y la técnica cumple este criterio.

Conjuntamente a los criterios Frye, Kelley y Daubert y a los requerimientos legales, procedimentales y científicos para la práctica forense, Arce (2016) creó una lista de requerimientos para su cumplimiento por las pruebas de credibilidad de testimonio. Bajo estas premisas, la Hipótesis Undeutsch y la técnica (SVA y CBCA) fueron sometidas a prueba:

1. ¿Es la teoría científica que subyace a la evidencia, válida, comprobable y validada en una publicación científica? La Hipótesis Undeutsch no se derivó originalmente de una teoría científica, sino de información proveniente de archivos de caso, esto es, de un procedimiento bottom-up (fue originaria de un elevado número de casos reales de abuso sexual infantil). No obstante, la teoría que subyace a la Hipótesis Undeutsch puede explicarse desde diferentes teorías de la memoria y el engaño como las teorías de la racionalidad y constructivistas las cuales establecen que quien miente tiene una visión estereotipada de la mentira, es decir, la mentira es planeada y aprendida y, por tanto, consistente en el tiempo. Así, las personas que mienten construyen las mentiras a partir de *scripts* cognitivos (Schank & Abelson, 1977). Mientras, teorías sobre la memoria episódica y autobiográfica apoyan la diferente calidad de las memorias basadas en hechos experimentados. El modelo del *reality monitoring* también

proporciona soporte teórico (las categorías de contenido de memoria discriminan entre memorias internas -imaginadas/pensadas- y externas -percibidas) a la hipótesis (los atributos de memoria discriminan entre ambas memorias) (Johnson & Raye, 1981). Finalmente, memorias sensoriales y narrativas pueden explicar las diferencias entre memorias auto-experimentadas y fabricadas, siendo los criterios de contenido el medio para ello. Los resultados de nuestros meta-análisis validan la hipótesis, esto es, la teoría i.e., *Hypothesis Undeutsch*, es sometida a prueba y fue probada y verificada a partir de las publicaciones científicas, es decir, las investigaciones provenientes de la comunidad científica concuerdan con su validez (ver criterios Daubert 1, 2, 4 y 5 para más detalles).

2. ¿La técnica que se deriva de la hipótesis es válida? La técnica que se deriva de la hipótesis, en definitiva, los criterios de realidad del CBCA, discriminan significativamente entre-contextos (e.g., menores, adultos, diferentes delitos, víctimas, testigos) entre memorias vividas y fabricadas. En consecuencia, la evidencia científica valida la técnica. Sin embargo, para la práctica forense no es válida porque no se proporciona un criterio de decisión estricto para la clasificación de las declaraciones, dejando que la decisión forense recaiga en una impresión del perito o juicio clínico. Por tanto, la decisión resulta semi-objetiva i.e., basada en evidencia científica general (objetiva) pero cuando se aplica a casos específicos la decisión se sustenta en la impresión del experto (subjetiva), mientras que los tribunales demandan un criterio de decisión objetivo.
3. ¿Se ha aplicado correctamente la técnica al caso particular? El SVA, como el sistema categorial global en el que se describe la aplicación de los criterios de realidad del CBCA, no cumple este criterio puesto que no hay una medida sobre el mismo. En consecuencia, los resultados no son fiables para la práctica forense ( $N = 1$ ).
4. ¿Los métodos utilizados pueden ser validados y revisados por otros expertos? Aunque el SVA no incluye este criterio, si las declaraciones son grabadas y guardadas, los métodos pueden ser validados.
5. ¿Pueden replicarse las pruebas (testimonio experto de un segundo perito)? Si la declaración del testigo puede ser obtenida una vez más, la prueba puede ser replicada.
6. ¿Puede evaluarse la consistencia del testimonio (test-retest)? Como las declaraciones obtenidas a partir de una técnica de entrevista no invasiva i.e., centrada en el recuerdo libre, pueden aplicarse en medidas repetidas (Memon, Meissner, & Fraser, 2010; Memon, Wark, Bull, & Köhnken, 1997), la consistencia del testimonio puede ser evaluada. Para ellos, Undeutsch creó una segunda hipótesis para las memorias de hechos vividos: los elementos centrales de la declaración son consistentes; mientras que los detalles periféricos pueden no serlo. Las teorías de recuperación de la memoria y la evidencia apoyan esta hipótesis (Fariña et al., 1994).
7. A pesar de que se conoce la tasa de error de clasificación (ver criterio Daubert 3, *¿Se conoce la tasa de error?*), no cumple el criterio jurisprudencial de presunción de inocencia (ninguna persona inocente puede estar en prisión, mientras que es suficiente con que una persona culpable esté generalmente en prisión; Sentencia

del TS 832/2000 de 28 de Febrero de 2000; Sentencia del TS 213/2002 de 14 de Febrero de 2002). Es decir, dado que la clasificación de una memoria fabricada como una memoria real implica una falsa acusación (falso positivo) en la evaluación forense esta tasa debe ser cero. La técnica falla en la clasificación de falsos positivos entre un 9.8% en menores y un 28.5% en adultos. En conclusión, la técnica no es válida para la práctica forense.

8. ¿Son las conclusiones sobre el testimonio válidas para el contexto judicial? El SVA proporciona 3 decisiones posibles basadas en: a) un juicio cualitativo guiado por reglas de decisión; b) una escala tipo Likert de 5 puntos, que va desde no creíble (-2) hasta creíble (+2), pasando por indeterminado (0) (Steller, 1989); y c) un juicio clínico i.e., cualitativo (Köhnken, 2004). Ninguna de las tres opciones son válidas para los tribunales por su falta de objetividad, replicabilidad y consistencia resultando, incluso, confusas (e.g., el Tribunal Supremo Español advierte sobre ello).
9. ¿Es válido el sistema de análisis de contenido? Las propiedades psicométricas del CBCA son desconocidas. Con motivo de que un sistema categorial sea una medida metódica (Bardin, 1995), las categorías de análisis deben ser mutuamente excluyentes (se han encontrado categorías que se superponen, provocando la duplicidad de medidas; Horowitz et al., 1997; Roma et al., 2011); homogéneas (se desconoce); exhaustivas (los criterios de realidad del CBCA no son exhaustivos puesto que pueden ser complementados con criterios de realidad adicionales, ver meta-análisis de adultos; Arce & Fariña, 2009); fidelidad (no se ha estimado la fidelidad del SVA en su aplicación al contexto forense; en la investigación científica, la fidelidad es variable); objetivas (las categorías no están definidas con precisión lo que deriva en diferentes interpretaciones y codificaciones); pertinencia (como el sistema ha sido formulado para memorias de menores que han sufrido abuso sexual, algunas categorías no son pertinentes en otros contextos de evaluación —ver meta-análisis de adultos— y, además, las estrategias de engaño dependen del contexto; Volbert & Steller, 2014); y productivas (los criterios de realidad del CBCA fueron productivos y significativos en la discriminación de memorias).
10. Cualificación del perito experto. Aunque el SVA no repara en la cualificación del experto, ciertos autores advierten que el entrenamiento en esta técnica es necesario y efectivo, tanto para el análisis de contenido como para la entrevista, ya que favorece la aparición de criterios de realidad (Fisher, Geiselman, & Raymond, 1987; Köhken, 2004; Vrij, 2005). En consecuencia, se requiere entrenamiento y, cuando sea necesario, una cualificación oficial como psicóloga o psicólogo forense (cuando la acreditación oficial no esté disponible puede sustituirse por una profesional).

La pericial psicológica, debe contemplar una medida de la cuantificación de secuelas además de la evaluación de la credibilidad del testimonio, consecuencia de la victimización de un abuso sexual en la infancia o adolescencia. Se ha estimado la elevada probabilidad de desarrollar un trastorno depresivo o de ansiedad en la adultez como consecuencia de la victimización (secuelas a largo plazo). En consecuencia, la cuantificación de la huella psicológica tras la victimización de un delito de naturaleza sexual, emerge como una forma de compensación económica hacia la víctima (así como la responsabilidad criminal) y esta

responsabilidad civil debe ser proporcional al daño causado. Aunque se debe estudiar cada caso individualmente, los resultados meta-analíticos con las variables moderadoras nos proporcionan una guía sobre el alcance del daño esperado de acuerdo a cada caso (i.e., sexo de la víctima, severidad del abuso). Puesto que el diagnóstico clínico emerge como una medida del daño significativamente más sensitiva que la medida de sintomatología, y la instrumentación psicométrica como tarea de reconocimiento favorece la simulación de sintomatología (en el contexto judicial debemos sospechar simulación de síntomas, DSM), se recomienda cuantificar el daño a través del diagnóstico clínico utilizando la entrevista ya que permite relacionar daño y victimización y controlar la simulación. La *entrevista clínico-forense* (Arce & Fariña, 2001) se presenta como una entrevista eficiente en la práctica forense para este cometido, puesto que favorece la aparición espontánea durante la declaración, de sintomatología relacionada con la victimización del delito (tarea de conocimiento).

En resumen, el principal objetivo de esta tesis doctoral fue proporcionar respuestas sobre la admisibilidad en los tribunales de sistemas categoriales para el análisis de la credibilidad. La técnica meta-analítica, con su labor integrativa de los resultados empíricos, permitió alcanzar conclusiones generales que contribuyen al enfoque de la Psicología Basada en la Evidencia. Los resultados apoyan la creación y validación de un sistema categorial metódico para la evaluación de la credibilidad, que debe estar compuesto por categorías de análisis de contenido válidas y fiables, y conformado por criterios de realidad específicos para cada contexto (i.e., delitos contra la libertad e indemnidad sexuales, violencia de género) ya que el contexto media en la discriminación de las categorías de realidad (Arce, Fariña, & Vilariño, 2010). De hecho, el CBCA no cumple algunos de los criterios legales (e.g., posee una elevada tasa de error de clasificación, falta de un criterio de decisión estricto) y requerimientos metodológicos (e.g., algunos criterios de realidad no son mutuamente excluyentes, y el sistema no es objetivo ni pertinente) en su forma actual. Asimismo, el método de evaluación debe permitir estudiar la victimización a partir de la medida de daño psicológico como consecuencia de un delito sexual, con el objetivo de reparar el daño causado proporcionalmente al sufrimiento padecido, incluso de forma retrospectiva (el daño puede aparecer después de un largo período de tiempo e incluso cronificarse). Los resultados de los meta-análisis proporcionan a los jurados y jueces una estimación del alcance del daño, arrojando luz sobre el elevado coste psicológico de la victimización de delitos de naturaleza sexual. Además, después del diagnóstico de un trastorno internalizante, se recomienda la aplicación de la Escala de Evaluación de la Actividad Global (EEAG) al caso concreto con el objetivo de cuantificar el nivel general de sufrimiento de la persona evaluada (ajuste psicológico en las diferentes facetas de su vida diaria). Esta medida permite reparar el daño proporcionalmente a las secuelas sufridas, y sería útil para ayudar a jueces en el alcance de sentencias judiciales más exactas y justas.

La investigación futura debe centrar sus esfuerzos en las siguientes líneas. Primero, mejorar la eficacia del sistema a través de la identificación de nuevas categorías productivas (proceso bottom-up en casos reales) en relación con contextos específicos. Asimismo, a partir de la combinación de los criterios de realidad del CBCA con el SRA, atributos de memoria (RM) y criterios adicionales procedentes de casos reales, la capacidad discriminativa del sistema categorial se espera que mejore. Es más, aunque la inclusión de “criterios de engaño” en el CBCA no había sido aceptado debido a su naturaleza, su inclusión puede ser un camino hacia la mejora del sistema (Steller, 1988). Segundo, los criterios del componente motivacional de la hipótesis, aunque fueron productivos, mostraron la menor capacidad

discriminativa. Estos criterios se relacionan con una asunción consciente de una estrategia de engaño, es decir, se evitan conscientemente en las declaraciones con el objetivo de parecer honestos. Por esta razón, es de relevancia estudiar dichos criterios en profundidad porque parecen estar influenciados en mayor medida por el contexto o situación de evaluación que los criterios cognitivos. Concretamente, los criterios ‘auto-desaprobación’ y ‘perdón al autor del delito’ deben ser validados en casos forenses ya que los hallazgos apuntan a que se trata de estrategias de *auto-presentación* y no criterios de contenido relacionados con la memoria, en consonancia con propuestas de otras autoras (Niehaus, 2008). Tercero, la investigación criminológica de los diferentes tipos de delitos sexuales ayudará, no solo a crear sistemas específicos para evaluar el testimonio, sino también para definir con mayor precisión el criterio “detalles característicos de la agresión” que su elevado efecto revela su importancia en el proceso de toma de decisión sobre la veracidad del testimonio. Cuarto, el CBCA discrimina entre memorias de eventos observados en vídeo (no auto-experimentados) y fabricados, aunque la hipótesis no ha sido formulada para tal propósito. Por tanto, a partir de estos hallazgos no hipotetizados originalmente, sería de gran interés analizar la capacidad del CBCA para discriminar entre declaraciones basadas en eventos auto-experimentados y observados en vídeo. Finalmente, se necesitan más estudios de campo con víctimas/testigos reales con el fin de obtener un criterio de decisión estricto para tomar decisiones más precisas sobre la realidad de un testimonio. Estudios experimentales bajo condiciones de alta fidelidad pueden lograr validez interna y contribuir significativamente al estudio de la credibilidad. Para ello, futuras investigaciones de laboratorio deben diseñarse bajo los siguientes criterios: la involucración del sujeto en los hechos, que éstos revistan un tono emocional negativo, que implique una pérdida de control sobre la situación (Steller, 1989), que la declaración se realice cara a cara y que la implicación o motivación del sujeto se controle, obtención de declaraciones largas no solo a partir del recuerdo libre sino también mediante preguntas no dirigidas ni sugestivas (reinstauración de contextos) y ser entrevistados en diferentes ocasiones (Volbert & Steller, 2014).

Los estudios meta-analíticos que pusieron a prueba la validez del CBCA tanto en muestras de menores como de adultos, comportan las mismas limitaciones y serán abordadas de forma conjunta. Primera, se desconoce la fiabilidad de las entrevistas previas al análisis de contenido. El pobre entrenamiento de los entrevistadores puede derivar en declaraciones de contenido insuficiente para un análisis de credibilidad o en testimonios orientados hacia respuestas “sugeridas”, como consecuencia del estilo de entrevista utilizado por el personal entrevistador (e.g., entrevistas con preguntas dirigidas y/o sugestivas). Del mismo modo, la mayoría de las investigaciones primarias no informaban del cumplimiento de los estándares del SVA, concretamente del tipo de entrevista utilizada, aspecto relevante para valorar de forma argumentada la calidad de la declaración. Estas consideraciones pueden favorecer el rechazo de la Hipótesis Undeutsch como resultado de unas malas prácticas. Segunda, los estudios experimentales han mostrado validez aparente (Konecni & Ebbesen, 1992), pero no validez ecológica o de generalización de resultados (Sarwar, Allwood, & Innes-Ker, 2014). El personal investigador en el contexto de laboratorio manipula las condiciones experimentales de modo que conocen la veracidad de las declaraciones que evalúan, esto es, la *ground truth*. Este hecho les lleva a utilizar el CBCA como una herramienta de detección de engaño, clasificando con certeza las declaraciones falsas. Sin embargo, en la práctica forense hay que evaluar otras hipótesis alternativas a la fabricación total y consciente del testimonio, como posible explicación a una declaración de baja calidad: falta de colaboración de la víctima, memoria insuficiente de los hechos para un análisis de contenido de la declaración, o capacidades cognitivas disminuidas (Köhnken, 2004). Tercera, la tendencia hacia la no

publicación de resultados estadísticamente no significativos y a la eliminación de categorías de realidad ineficaces en la discriminación de memorias reales e inventadas (*conflicting data*), favorece que los resultados hallados tomen la dirección de la argumentación de la Hipótesis Undeutsch. Cuarta, la fiabilidad de los criterios de realidad fue una estimación puesto que muchas de las investigaciones primarias no reportaban el acuerdo entre evaluadores. Igualmente, la fiabilidad calculada únicamente a partir de la frecuencia de aparición de los criterios sin tener en cuenta su exacta correspondencia, sobreestima la fiabilidad de éstos. Además, la fiabilidad en el análisis de contenido se deriva, no solo de la consistencia entre codificadores, sino también de la consistencia intra-codificador (fiabilidad en el tiempo) e inter-contextos (consistencia con otros evaluadores distintos a los del estudio y otros materiales). Una codificación más fiable anularía los potenciales efectos del sesgo de realidad (truth bias) o el sesgo de respuesta que se ha asociado con la aplicación de los criterios de realidad (Griesel, Ternes, Schraml, Cooper, & Yuille, 2013; Rassin, 1999; Sporer, 2004). Por último, algunos resultados meta-analíticos pueden estar sujetos a cierta variabilidad cuando  $N_s < 400$ , no garantizando la estabilidad en las estimaciones de muestreo (Hunter & Schmidt, 2015).

De otra parte, las limitaciones halladas en el meta-análisis del efecto de la victimización de abusos sexuales en el padecimiento de secuelas psicológicas fueron las que siguen. Primero, al tratarse de investigaciones sobre casos reales de abusos sexuales en la infancia y adolescencia, la *ground truth* descansa en medidas auto-informadas y, en numerosas ocasiones, de naturaleza retrospectiva. Esto dificulta la comprobación de la veracidad de los hechos, infraestima la relación entre victimización y daño, y facilita la comisión de errores de clasificación, especialmente de falsos negativos. Segunda, se ha asumido la sintomatología y diagnóstico de un trastorno depresivo y/o de ansiedad como consecuencia de la victimización de un acto de violencia sexual, sin embargo, la relación causa-efecto entre los hechos y la huella psíquica no se había establecido explícitamente. Tercera, algunas investigaciones primarias comportaban una medida de abuso sin diferenciar entre las distintas tipologías (e.g., abuso sexual, abuso físico, negligencia). Cuarta, en tanto que determinados estudios no reportaban un grupo control, se utilizó la población normativa como grupo de contraste.

## 6. DISCUSSION

From meta-analytic results, the following conclusions have arisen:

### **In children memories:**

1. The Undeutsch Hypothesis was validated in child memories through the study of CBCA reality criteria. That is, CBCA as a categorical system for analyzing the content of children statements discriminated significantly between vivid and fabricated memories. This question leads to a major number of reality criteria in statements based on real events rather than in fabricated or invented statements. Moreover, results supported Undeutsch Hypothesis for all children ages and contexts (i.e., for memory contexts other than sexual abuse).
2. The total CBCA score discerned between real and fabricated statements, independently from the research paradigm (experimental studies vs. field studies). However, the system's efficacy was significantly superior in field rather than in experimental studies. In consequence, experimental studies results are not directly generalizable to practice unless they are supported by field studies.
3. All CBCA reality criteria discriminated significantly between vivid and fabricated memories, that is, results validated all CBCA reality criteria.
4. Results revealed greater effects sizes for cognitive reality criteria (CBCA criteria 1 to 13) than for motivational ones (14 to 18). Besides, motivational criteria were non-generalizable.
5. The 'general characteristics' major category i.e., 'logical structure', 'unstructured production', 'quantity of details' criteria, have reached the greatest effect sizes. The largest effect size was for the 'details characteristic of the offence' criterion. We suggest that this may rest on the difficulty to fabricate these kind of details.
6. As for forensic practice, the total CBCA score classifies correctly 68.5% of the self-experienced memories (true positives), meanwhile fails in the detection of 31.5% self-experienced memories (false negatives). As for true negatives, results may not be generalized to practice as reality criteria do not classify false memories. Regarding false positives (false statements identified as truthful), that in forensic settings they must be zero as otherwise it would infringe the principle of the presumption of innocence (i.e., the Spanish Constitutional and the Supreme Courts have stated that no innocent person should be imprisoned, while some guilty person may be out of jail), the potential probability is .288 (1-.712, that is, the probability of getting more CBCA criteria in children testimonies of vivid events).

7. The analysis of the research paradigm moderator showed a greater effect size and true positive classification accuracy of the total CBCA score for field studies (90.2%) than for experimental studies (63.5%). In consequence, results from experimental studies may not be generalized directly for practice, revealing face validity for experimental data (Fariña et al., 1994; Konecni & Ebbesen, 1992). This also corrects the false positive rate passing to .029. Nevertheless, the strict criterion for this remains unknown, which means that this procedure is not a valid/power proof since it is based on a global impression or a clinical judgment (Arce & Fariña, 2013; Köhnken, 2004).

### **In adult memories:**

8. Additionally, the Undeutsch Hypothesis was tested for adult population (since Undeutsch Hypothesis was grounded on memory content, it was already theoretically anticipated that the hypothesis would be also appropriate both in adults and in contexts other than child sexual abuse) (Berliner & Conte, 1993). Likewise, Undeutsch Hypothesis was validated not only for victims/claimants, witnesses and offenders' memories but for experimental (high internal validity) and field (high external validity) studies. However, CBCA reality criteria classified vivid memories significantly superior in field rather than in simulation study designs, not being directly generalized by their own to practice.
9. Not all reality criteria were valid (i.e., 'self-deprecation', 'pardoning the perpetrator') nor generalizable (i.e., all criteria with the exception of 'accurately reported details misunderstood' and 'details characteristic of the offence' which are affected by a second-order sampling error and it is not possible to do an estimation). Notwithstanding, 'pardoning the perpetrator' criterion performed against Undeutsch Hypothesis (negative mean effect size).
10. Even though Undeutsch Hypothesis and reality criteria are also valid for offenders' memories, results may not be undoubtedly extrapolated to practice as data has exclusively come from experimental studies. Even so, real offender memories in forensic evaluations have not a ground truth because offenders are able, by law, to lie, to not declare, or they simply may deny the alleged event alternatively to telling the truth. Under these considerations, the 'logical structure', 'unstructured production', 'quantity of details', 'contextual embedding', "descriptions of interactions", 'reproduction of conversations', 'unexpected complications', 'superfluous details', 'perpetrators mental state', 'admitting lack of memory' and 'details characteristic of the offence' criteria distinguished significantly between the offender's vivid and fabricated memories, being the last two criteria, as well as the total CBCA score, generalized to any condition.
11. Results support the additional reality criteria to those involving CBCA reality criteria (i.e., 'reporting style', 'display insecurities', 'providing reasons for lack of memory'). In consequence, CBCA criteria may be supplemented with additional reality criteria.



12. Some of the studied additional reality criteria are not in fact reality criteria, as they are inversely related to vivid memories. Thus, 'repetitions' and 'clichés' criteria have a negative relation with vivid memories. Therefore, these are memory attributes i.e., memory characteristics not directly related to reality, that should be studied to create a categorical system combining reality criteria and memory attributes. On this, Arce and Fariña (2009; Vilariño et al., 2011) created and validated a categorical system to discriminate vivid and fabricated memories of intimate partner violence claimants.
13. Overall, field studies exhibited that not all reality criteria are valid to discriminate between true and fabricated memories. These results suggest reducing the categorical system to significant discriminant criteria, as the non-significant criteria only introduces noise.
14. Consequently, the moderator variable 'evaluation setting' revealed the best performance of the reality criteria for statements of sexual offences and intimate partner violence field studies, classifying 71.5% of true positives correctly and failing in 28.5% of false positives.
15. The efficiency of the total CBCA score to distinguish between fabricated and vivid memories was significantly higher for children than for adults.
16. Comparatively, 'contextual embedding', 'superfluous details', 'subjective mental state' and 'details characteristic of the offence' criteria enhanced classification rates in children. Remaining criteria performed equally in adults and children.

### **Child and adolescent sexual abuse internalizing symptomatology**

17. Child or adolescent sexual abuse victimization (CSA/ASA) implies approximately a probability of .70 of suffering internalizing injury, of .66 of depression and of .68 of anxiety. Damage quantification was estimated at about 30% in general sequelae, depression, and anxiety.
18. In relation to the diagnosis of internalizing disorders, female victims have 9% more probability of developing depression than male victims, but not anxiety. This derives in large damage in depression (41 vs. 10%, for females and males, respectively) and anxiety (24 vs. 14%) for females than for males. This result is linked to epidemiological studies findings and international organizations classifying mental disorders prevalence informed (APA, 2013; WHO, 2000) pointing out the vulnerability of women in developing internalizing sequelae.
19. The probability of chronic injury (dysthymia) was significantly greater than more severe injury (major depressive disorder).
20. Damage quantification for a persistent depressive disorder (dysthymia) and for a major depressive disorder was about 46 and 31% for CSA/ASA victims, respectively.

21. CSA/ASA victimization increases 43% the probability of developing an anxiety disorder (generalized anxiety disorder, specific phobia, social phobia, or panic disorder).
22. In short, suffering CSA/ASA predisposes victims to psychological adjustment problems in adulthood (Chiesa, Larsen-Paya, Martino, & Trinchieri, 2016).
23. Clinical diagnosis is a significantly more adequate (sensitive) injury measure than symptomatology for CSA/ASA victims. In consequence, clinical diagnosis should be preferred as a benchmark, rather than symptoms to assess the CSA/ASA victimization sequelae.
24. Severe sexual aggression victimization (i.e., with penetration), in contrast to non-contact aggression, causes a higher anxious and depressive injury. This finding gives support to Criminal Codes adjudicating more severe punishment when the sexual crime implies a penetration, as it causes a greater psychological damage than sexual crimes with non-contact aggression.

### **Implications of the results for forensic practice**

Admissibility of scientific evidence in judicial setting rests on meeting standards established by courts and the law of precedent. On this, the answers to such standards about the admissibility of the CBCA reality criteria throughout meta-analyses results are presented below:

1. Is the scientific hypothesis testable? The underlying hypothesis of the CBCA categorical system, the Undeutsch Hypothesis, has been submitted to proof in many studies (all the studies included in the first, second and fourth meta-analyses involved in the present doctoral dissertation submitted to test the hypothesis). Thus, the underlying scientific hypothesis may and has been submitted to contrast.
2. Has the hypothesis been tested? The hypothesis assessed through the CBCA has been tested and firmly confirmed (on the basis of our meta-analyses).
3. Is there a known error rate? Originally, the error rate was neither provided, nor considered (Köhnken, 2004; Steller, 1989; Steller & Köhnken, 1989; Undeutsch, 1967, 1988) as the resulting technique was semi-objective or based on expert clinical judgments. Vrij (2005, 2008) estimated (qualitative study) the error rate around 30% for true and false statements. Nevertheless, the results for false memories are not valid as the hypothesis and criteria do not classify them. Our results stated different error rates in truthful memories classification (false positives), being the most noteworthy, 9.8% in real child sexual abuse cases and 28.5% in real sexual and intimate partner violence adult cases.
4. Has the hypothesis/technique been subjected to peer review publication? Yes, many studies contrasting the hypothesis and technique have been published. Most of them submitted for peer review.

5. Is the theory upon which the hypothesis and/or technique is based on generally accepted among appropriate scientific community? Although the scientific community has not been consulted as such, scientific evidence coming from the appropriate researchers give support (significant and generalized in field context; see the results of our meta-analyses) to the hypothesis and technique. As this is not a minimal support, in which case results should be considered with skepticism (*Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 1993), the reasonable conclusion is that the hypothesis and technique meet this standard.

Joining the Frye, Kelley and Daubert Standards with scientific, procedural and legal requirements to forensic proofs, Arce (2016) created a list of testimony credibility proof requirements for which the Undeutsch Hypothesis and technique (SVA and CBCA) are submitted:

1. Is the scientific theory, upon which the evidence-based is valid, testable and tested in a scientific publication? The Undeutsch Hypothesis was not originally derived from a scientific theory, but from data, that is, bottom-up (it was originated from an elevated number of real-life child sexual abuse criminal cases). However, the underpinning Undeutsch Hypothesis theory can be explained by different memory and deception theories such as rationality and constructivism which state that liars have a stereotypical notion of lies, that is, a lie is planned and learned and consequently consistent in time. Thus, liars construct their lies from cognitive scripts (Schank & Abelson, 1977). Meanwhile, theories about episodic and autobiographical memories support different quality of memories from self-experienced events. The reality monitoring model also gives theoretical support (memory content categories discriminate between internal —imagined/thought— and external —perceived— memories) to the hypothesis (memory attributes discriminate between both memories) (Johnson & Raye, 1981). Finally, sensorial and narrative memories may explain the differences between self-experienced and fabricated memories, being the content criteria the media for this. The results of our meta-analyses validate the hypothesis, that is, the theory i.e., Undeutsch Hypothesis, is testable and was tested and verified by the scientific publications, that is, the scientific community research agrees on its validity (see Daubert standards 1, 2, 4 and 5 for more details).
2. Is the technique derived from the hypothesis valid? The technique derived from the hypothesis, the reality criteria of the CBCA, discriminate significantly between-contexts (e.g., children, adults, different crimes, victims, witnesses) between vivid and fabricated memories. In consequence, scientific evidence validates the technique. Nonetheless, for forensic practice it is not valid because a strict decision criterion to classify statements is not provided, leaving the forensic decision resting on the expert impression or clinical judgment. Thus, the resulting decision is semi-objective i.e., based on general scientific evidence (objective) but when applied to a specific case sustained in the expert impression (subjective), while an objective decision criterion is expected in court.
3. Has been the technique applied correctly in a particular case? The SVA, as the whole procedural system where the application of the CBCA reality criteria is

described, does not meet this standard as it does not have any measure. In consequence, forensic practice results ( $N = 1$ ) are not reliable.

4. Is it possible for the methods to be testable and revised by other experts? Although the SVA does not include this standard, if the statements are recorded and maintained, the methods may be tested.
5. Is it possible for the evidence to be replicated (second expert testimony)? If the witness statement could be obtained once again, the evidence may be replicated.
6. Is it possible for the persistence in the testimony (test-retest) to be assessed? As the statements obtained with non-invasive interview techniques i.e., centered in a free narrative recall, may be applied in repeated measures (Memon, Meissner, & Fraser, 2010; Memon, Wark, Bull, & Köhnken, 1997), testimony persistence may be assessed. As for this, Undeutsch created a second hypothesis: the central elements of the account are persistent; meanwhile the peripheral ones may not be (expected results for vivid memories). Memory recall theories and evidence support this (Fariña, Arce, & Real, 1994).
7. Meanwhile, the error rate is known (*Is the error rate known?* Daubert standard 3, see above), it does not meet the law of precedent standard of the presumption of innocence (no one innocent person shall be imprisoned, whereas it is enough that a guilty person generally stays imprisoned; 832/2000 28th February, 2000 Spanish Supreme Court Sentence; 213/2002 14th February, 2002 Spanish Supreme Court Sentence). That is, given that the classification of a fabricated memory as a real memory involves a false accusation (false positive), in forensic evaluation its rate must be zero. The technique fails in the classification of false positives among 9.8% in children and 28.5% in adults. In consequence, the technique is not valid for forensic purposes.
8. Are the conclusions about the testimony valid in a judicial setting? SVA establishes three decision options: a) in a qualitative judgment guided by rules; b) in a five-point Likert scale, ranging from incredible (-2) to credible (+2), passing through indeterminate (0) (Steller, 1989); and c) in an expert clinical judgement i.e., qualitative (Köhnken 2004). These three options are not valid for courts as they lack objectivity, replicability, and consistency, and are even confusing for courts (e.g., Spanish Supreme Court advertises about this).
9. Is the content analysis system valid? Psychometric properties of the CBCA are unknown. In order for a categorical system to be a methodic measure (Bardin, 1977), category analysis must be mutually exclusive (overlapping between categories has been found, producing duplicity of measures; Horowitz et al., 1997; Roma et al., 2011); homogeneous (unknown); exhaustive (CBCA reality criteria system is not exhaustive as additional criteria may be added, see adult meta-analysis; Arce & Fariña, 2009); fidelity (SVA has not an estimation of the fidelity of its application in forensic setting; in scientific research, fidelity is variable); objective (categories are not defined with precision giving to different interpretations and codings); pertinence (as the system was formulated for child sexual abuse memories, some categories are not pertinent in other evaluation contexts —see adult meta-analysis—, and deception strategies depend on the

context; Volbert & Steller, 2014); and productive (all the CBCA reality criteria were productive and significant to discriminate memories).

10. Expert witness qualification. Although SVA does not realize the expert qualification, authors advertise that training in this technique is necessary and effective, in both content analysis and interview techniques, eliciting reality criteria (Fisher, Geiselman, & Raymond, 1987; Köhken, 2004; Vrij, 2005). In consequence, training is required and, when available, an official qualification as a forensic psychologist (professional accreditation may substitute official if it is not available).

Expert psychological evidence must contemplate sequelae quantification measures in addition to the testimony credibility assessment, consequence of child sexual abuse or adolescent sexual abuse victimization. It has been estimated that there is an elevated probability of developing depression and anxiety disorders in adulthood as a consequence of victimization (long-term sequelae). Consequently, quantifying psychological injury from sexual crime victimization emerges as a way to compensate victim economically (as well as crime responsibility) and this civil responsibility must be proportional to the injury caused. Even though we must study each isolated case, meta-analytic results with moderator variables give us a guide to the extent of the injury expected according to each case (i.e., sex of the victim, severity of the abuse). As clinical diagnosis emerges as a significantly more sensitive injury measure than the symptoms, and psychometric instruments, as a recognition task, they favor symptomatology malingering (in the judicial setting we must suspect malingering, DSM). Measuring damage from clinical diagnosis through an interview is recommended because it allows both linking injury and victimization and to control malingering. *Forensic-clinical interview* (Arce & Fariña, 2001) has arisen as an efficient interview in forensic practice for this issue because it elicits spontaneous symptomatology reports related to the crime (knowledge task).

In short, the main objective of this doctoral dissertation was to provide answers to the concerns about the admissibility in court of the credibility assessment categorical system. Meta-analytic technique with its integrative labor of empirical results permitted reaching generalizable conclusions that contribute to the Evidence-Based Psychology (BEP) approach. Findings found the creation and validation of the methodic credibility assessment categorical system, which should be made up of reliable and valid categories of content analysis, conformed by significant reality criteria and contextually specific (i.e., sexual violence, intimate partner violence). In fact, CBCA does not fulfil some of these legal (e.g., elevated classification error rate, lack of a strict decision criterion) and methodological requirements (e.g., some reality criteria are non-mutually exclusive, and the system is neither objective nor pertinent) in its present state. Likewise, the assessment method must allow studying psychological injury victimization and its measure as a consequence of a sexual crime, with the objective of repairing damage proportionally to the distress suffered even retrospectively (injury may appear after a long period of time and even become chronic). Meta-analysis results provide juries and judges a prospectively estimated measure of injury, which shed some light the extent of the cost of psychological damage in sexual crimes victimization. In addition, after diagnosing an internalizing disorder, Global Assessment of Functioning (GAF) scale must be applied to the particular case in order to quantify the individual's overall level of distress (psychological adjustment in different life facets). This measure allows repair

damage proportionally to the distress suffered, and it would be helpful for judges to reach a more accurate and fair judicial sentence.

Future research must focus their efforts on the following investigation lines. First, improving system effectiveness through identifying new productive categories (real cases bottom-up process), regarding specific contexts. Likewise, combining CBCA reality criteria with SRA criteria, memory attributes (RM) and additional criteria coming from real-cases, the categorical system is expected to improve its discriminative capacity. Even more, while the inclusion of “deception criteria” in CBCA had not been accepted because of its original nature, according to Steller (1988), their integration might be a path to a better performance of the categorical system. Second, motivational criteria showed the lowest discriminative capacity but they were nonetheless productive. They are related to a mindful deception strategy, deceivers avoid consciously motivational criteria in his or her statements in order to appear honest. For this reason, it remains relevant to study deeply those criteria, because they seem to be more influenced by context or situation assessment than criteria, which are part of the cognitive component of the hypothesis. Concretely, ‘self-deprecation’ and ‘pardoning the perpetrator’ must be tested in real forensic cases because findings have pointed out that they seem not to be memory-related content criteria but self-presentation strategies, consistent with other proposals (Niehaus, 2008). Third, research in criminological characteristics of different type of sexual offenses will help not only to create a specific testimony assessment protocol but also to define ‘details characteristics of the offence’ criterion that could be highly relevant in a truthful decision-making procedure. Fourth, CBCA differentiates between memories of video-observed events (non-self-experienced) and fabricated events, even though the hypothesis has not been formulated for that aim. Accordingly, with these originally non-hypothesized findings, it would be of great interest to analyze the CBCA capacity to discriminate between statements based on self-experienced events and video-observed events. Finally, more field studies with real victims/witnesses should be performed in order to obtain a strict decision criterion to make an accurate decision about the reality of a given testimony. Experimental studies designed under high fidelity conditions may engage internal validity and contribute significantly to research in credibility. Therefore, future laboratory research might include the following criteria: personal involvement of the participant in the scenario, negative emotional tone of an event, loss of control over the situation (Steller, 1989), and giving a testimony in a face-to-face situation under which the implication or motivation of the participants is controlled, eliciting long statements not only by a free recall but asking questions, and being questioned repeatedly (Volbert & Steller, 2014).

The meta-analytic studies that have tested CBCA validity in both children and adults’ samples share certain limitations that must be taken into account. First, interviews’ reliability prior to the content analysis is unknown. A poor interviewers training may lead to obtain insufficient statements for conducting a credibility analysis or to get biased testimonies as a result of suggestive questions stemmed from the interviewer’s style (e.g., leading and/or suggestive questioning style). Moreover, most of the primary investigations did not inform about the SVA standards compliance. Concretely, there was a lack of information regarding the type of interview used, which is a relevant aspect for the statement’s quality evaluation. These considerations may favor Undeutsch Hypothesis rejection as a result of bad practices. Second, experimental studies had no proven external validity (Sarwar, Allwood, & Innes-Ker, 2014), but only ‘face validity’ (Konecni & Ebbesen, 1992). Researchers in laboratory context manipulate experimental conditions in order to know the veracity of statements, that is, the

ground truth. This leads them to the use of the CBCA as a deception detection tool, classifying false statements with confidence. Nevertheless, in the forensic practice, other alternative hypotheses to the total and conscious fabrication of the testimony must be considered as a possible explanation for the poor quality of the statement: lack of victims' collaboration, insufficient memories regarding the incident to analyze the content of the statement, or diminished cognitive capacities (Köhnken, 2004). Third, the trend towards to not publish non-significant results and to remove non-efficient reality categories to discriminate between real and fabricated memories (conflicting data), favors Undeutsch Hypothesis fulfillment. Fourth, reality criteria reliability was estimated since most of the primary studies did not report the raters' consistency. Additionally, the estimated reliability based on the presence of the reality criteria frequency without identifying the exact correspondence on statement overestimates their reliability. Furthermore, content analysis reliability is related not only to coders' consistency but also to within-coder consistency (reliability in time) and between-contexts consistency (with other raters independent to the study, and other materials). Conducting a more reliable codification would offset the potential effect of a truth bias or a response bias associated with the application of reality criteria (Griesel, Ternes, Schraml, Cooper, & Yuille, 2013; Rassin, 1999; Sporer, 2004). Finally, some meta-analytic results may be subject to certain variability when  $N_s < 400$  so they do not guarantee the stability of sampling estimates (Hunter & Schmidt, 2015).

Several limitations were also found in the meta-analytic studies of the psychological consequences of sexual abuse victimization. First, since primary investigations were real cases of child and adolescent sexual abuse, the ground truth rests on self-reports of a retrospective nature in most cases. This aspect hinders veracity validation of facts, underestimates the relationship between victimization and injury and facilitates classification errors, especially false negatives. Second, the symptomatology and the diagnosis of a depressive and/or anxiety disorder as a consequence of a victimization of a sexual crime was assumed, however, cause-effect relation between facts and psychological damage has not been explicitly established. Third, some primary investigations informed about an abuse measure without differentiating between typologies (e.g., sexual abuse, physical abuse, negligence). Fourth, since some studies did not report a control group, normative population was taken as the contrast group.





## 7. REFERENCIAS

- American Psychiatric Association. (APA, 2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: American Psychiatric Association.
- Arce, R. (2016). *Evaluación del SVA/CBCA y el SEG: Criterios Frye, Kelly, Daubert, jurisprudenciales, procesales y legales*. Manuscrito Inédito. Santiago de Compostela, España: Unidad de Psicología Forense.
- Arce, R., & Fariña, F. (2001). *Construcción y validación de un procedimiento basado en una tarea de reconocimiento para la medida de la huella psíquica en víctimas de delitos: La entrevista forense*. Manuscrito inédito. Universidad de Santiago de Compostela.
- Arce, R., & Fariña, F. (2005). Psychological evidence in court on statement credibility, psychological injury and malingering: The Global Evaluation System (GES). *Papeles del Psicólogo*, 26, 59-77. Retrieved from <http://www.papelesdelpsicologo.es/English/1247.pdf>
- Arce, R., & Fariña, F. (2006). Psicología del testimonio: Evaluación de la credibilidad y de la huella psíquica en el contexto penal. In Consejo General del Poder Judicial (Ed.), *Psicología del testimonio y prueba pericial* (pp. 39-103). Madrid: Consejo General del Poder Judicial.
- Arce, R., & Fariña, F. (2009). Evaluación psicológica forense de la credibilidad y daño psíquico en casos de violencia de género mediante el Sistema de Evaluación Global. In F. Fariña, R. Arce, & G. Buela-Casal (Eds.), *Violencia de género. Tratado psicológico y legal* (pp. 147-168). Madrid: Biblioteca Nueva.
- Arce, R., & Fariña, F. (2013). Psicología forense experimental. Testigos y testimonio. Evaluación cognitiva de la veracidad de testimonios y declaraciones. In S. Delgado (Dir. Tratado), S. Delgado, & J. M. Maza (Coords. Vol.), *Tratado de medicina legal y ciencias forenses: Vol. V. Psiquiatría legal y forense* (pp. 21-46). Barcelona, España: Bosch.
- Arce, R., & Fariña, F. (2015). Evaluación psicológico-forense de la credibilidad y daño psíquico mediante el Sistema de Evaluación Global. In P. Rivas & G. L. Barrios (Dirs.), *Violencia de género: Perspectiva multidisciplinar y práctica forense* (pp. 657-367). Navarra, España: Thomson Aranzadi.
- Arce, R., Fariña, F., & Vilariño, M. (2010). Contraste de la efectividad del CBCA en la evaluación de la credibilidad en casos de violencia de género [Contrasting the efficiency of the CBCA in the assessment of credibility in violence against women cases]. *Intervención Psicosocial*, 19, 109-119. doi: 10.5093/in2010v19n2a2

- Arce, R., Fariña, F., & Vilariño, M. (2015). Daño psicológico en casos de víctimas de violencia de género: Un estudio comparativo de las evaluaciones forenses [Psychological injury in intimate partner violence cases: A contrastive analysis of forensic measures]. *Revista Iberoamericana de Psicología y Salud*, 6, 72-80. doi: 10.1016/j.rips.2015.04.002
- Arce, R., Seijo, A., & Novo, M. (2009). Validez del testimonio: Un estudio comparativo de los criterios legales y empíricos [Testimony validity: A comparative study of legal and empirical criteria]. *Anuario de Psicología Jurídica*, 19, 5-13. Retrieved from <http://www.redalyc.org/pdf/3150/315025340002.pdf>
- Baker, E. V. (1990). Psychological expert testimony on a child's veracity in child sexual abuse prosecutions. *Family Law Symposium*, 50, 1039-1055.
- Bardin, L. (1977). *L'Analyse de contenu* [Content analysis]. Paris, France: Presses Universitaires de France.
- Berliner, L., & Conte, J. R. (1993). Sexual abuse evaluation: Conceptual and empirical obstacles. *Child Abuse & Neglect*, 17, 111-125. doi: 10.1016/0145-2134(93)90012-T
- Blanchard, E. B., Hickling, E. J., Freidenberger, B. M., Malta, L. S., Kuhn, E., & Sykes, M. A. (2004). Two studies of psychiatric morbidity among motor vehicle accident survivors 1 year after the crash. *Behaviour Research and Therapy*, 42, 569-583. doi: 10.1016/S0005-7967(03)00162-1
- Bogaard, G., Meijer, E. H., Vrij, A., & Merckelbach, H. (2016). Scientific Content Analysis (SCAN) cannot distinguish between truthful and fabricated accounts of a negative event. *Frontiers in Psychology*, 7, 1-7. doi: 10.3389/fpsyg.2016.00243
- Botella, J., & Gambara, H. (2002). *Qué es el meta-análisis*. Madrid: Biblioteca Nueva.
- Briere, J. (1992). Methodological issues in the study of sexual abuse effects. *Journal of Consulting and Clinical Psychology*, 60, 196-203. doi: 10.1037/0022-006X.60.2.196
- Chiesa, M., Larsen-Paya, M., Martino, M., & Trinchieri, M. (2016). The relationship between childhood adversity, psychiatric disorder and clinical severity: Results from a multi-centre study. *Psychoanalytic Psychotherapy*, 30, 79-95. doi: 10.1080/02668734.2016.1145131
- Cohen, J. (1988). *Statistical power analysis for behavioral sciences* (2nd ed.). New York, NY: Academic Press.
- Daubert v. Merrell Dow Pharmaceuticals, 509 U.S. 579 (1993).
- De Luca, S., Navarro, F., & Cameriere, R. (2013). La prueba pericial y su valoración en el ámbito judicial español. *Revista Electrónica de Ciencia Penal y Criminología*, 15, 1-19.
- DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological Bulletin*, 1, 74-118. doi: 10.1037/0033-2909.129.1.74

- Echeburúa, E., Corral, P., & Amor, P. J. (2002). Evaluación del daño psicológico en las víctimas de delitos violentos. *Psicothema*, *14*, 139-146. Retrieved from <http://www.psicothema.com/pdf/3484.pdf>
- Echeburúa, E., Corral, P., Zubizarreta, I., & Sarasúa, B. (1995). *Trastorno de estrés postraumático crónico en víctimas de agresiones sexuales*. A Coruña: Fundación Paidea.
- Fariña, F., Arce, R., & Real, S. (1994). Ruedas de identificación: De la simulación y la realidad [Lineups: A comparison of high-fidelity research and research in a real context]. *Psicothema*, *6*, 395-402. Retrieved from <http://www.psicothema.es/pdf/935.pdf>
- Fariña, F., Arce, R., Vilariño, M., & Novo, M. (2014). Assessment of the standard forensic procedure for the evaluation of psychological injury in intimate-partner violence. *The Spanish Journal of Psychology*, *17*, e32, 1-10. doi: 10.1017/sjp.2014.30
- Fergusson, D. M., McLeod, G. F. H., & Horwood, L. J. (2013). Childhood sexual abuse and adult developmental outcomes: Findings from a 30-year longitudinal study in New Zealand. *Child Abuse & Neglect*, *37*, 664-674. doi 10.1016/j.chiabu.2013.03.013
- Fisher, R. P., & Geiselman, R. E. (1992). *Memory enhancing techniques for investigative interviewing: The cognitive interview*. Springfield, IL: Charles C. Thomas.
- Fisher, R. P., Geiselman, R. E., & Raymond, D. S. (1987). Critical analysis of police interviewing techniques. *Journal of Police Sciences and Administration*, *15*, 177-185.
- Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: Current use, calculations, and interpretation. *Journal of Experimental Psychology: General*, *141*, 2-18. doi: 10.1037/a0024338
- Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Education & Educational Research*, *5*, 3-8. doi: 10.3102/0013189X005010003
- Gold, L. H. (2014). DSM-5 and the assessment of functioning: The World Health Organization Disability Assessment Schedule 2.0 (WHODAS 2.0). *Journal of the American Academy of Psychiatry and the Law*, *42*, 173-181.
- Griesel, D., Ternes, M., Schraml, D., Cooper, B. S., & Yuille, J. C. (2013). The ABC's of CBCA: Verbal credibility assessment in practice. In B. S., Cooper, D. Griesel, & M. Ternes (Eds.), *Applied issues in investigative interviewing, eyewitness memory, and credibility assessment* (pp. 293-323). New York, NY: Springer. doi: 10.1007/978-1-4614-5547-9\_12
- Griesel, D., & Yuille, J. C. (2007). Credibility assessment in eyewitness memory. In M P. Toglia, J. D. Read, D. F. Ross, & R.C.L. Lindsay (Eds.), *Handbook of eyewitness psychology. Volume 1. Memory for events* (pp. 339-370). New York, NY: Lawrence Erlbaum Associates.

- Hardt, J., & Rutter, M. (2004). Validity of adult retrospective reports of adverse childhood experiences: Review of the evidence. *Journal of Child Psychology and Psychiatry*, *45*, 260-273. doi: 10.1111/j.1469-7610.2004.00218.x
- Hauch, V., Sporer, S. L., Michael, S. W., & Meissner, C. A. (2014). Does training improve the detection of deception? A meta-analysis. *Communication Research*, *43*, 283-343. doi: 10.1177/0093650214534974
- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.
- Hillberg, T., Hamilton-Giachritsis, C., & Dixon, L. (2011). Review of meta-analyses on the association between child sexual abuse and adult mental health difficulties: A systematic approach. *Trauma, Violence & Abuse*, *12*, 38-49. doi: 10.1177/1524838010386812
- Horowitz, S. W. (1991). Empirical support for Statement Validity Assessment. *Behavioral Assessment*, *13*, 293-313.
- Horowitz, S. W., Lamb, M. E., Esplin, P. W., Boychuk, T. D., Krispin, O., & Reiter-Lavery, L. (1997). Reliability of criteria-based content analysis of child witness statements. *Legal and Criminological Psychology*, *2*, 11-21. doi: 10.1111/j.2044-8333.1997.tb00329.x
- Hunter, J. E., & Schmidt, F. L. (2015). *Methods of meta-analysis: Correcting error and bias in research findings*. Newbury Park, CA: Sage.
- Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological Review*, *88*, 67-85.
- Köhnken, G. (2004). Statement Validity Analysis and the detection of the truth. In P. A. Granhag & L. A. Strömwall (Eds.), *The detection of deception in forensic contexts* (pp. 41-63). Cambridge, UK: Cambridge University Press. doi: 10.1017/CBO9780511490071.003
- Köhnken, G., Schimossek, E., Aschermann, E., & Höfer, E. (1995). The cognitive interview and the assessment of the credibility of adults' statements. *Journal of Applied Psychology*, *80*, 671-684. doi:10.1037/0021-9010.80.6.671
- Köhnken, G., & Steller, M. (1988). The evaluation of the credibility of child witness statements in the German procedural system. *Issues in Criminological and Legal Psychology*, *13*, 37-45.
- Konecni, V. J., & Ebbesen, E. B. (1992). Methodological issues on legal decision-making, with special reference to experimental simulations. In F. Lösel, D. Bender, & T. Bliesener (Eds.), *Psychology and law. International perspectives* (pp. 413-423). Berlin, Germany: Walter de Gruyter.
- Manzanero, A. L., & Muñoz, J. M. (2011). *La prueba pericial psicológica sobre la credibilidad del testimonio. Reflexiones psico-legales*. Madrid, España: SEPIN.

- McGough, L. S. (1991). Commentary, Assessing the Credibility of Witnesses' Statements. In J. L. Doris (Ed.), *The Suggestibility of Children's Recollections* (pp. 165-167). Washington, DC: American Psychological Association.
- McGraw, K. O., & Wong, S. P. (1992). A common language effect size statistic. *Psychological Bulletin*, *111*, 361-365.
- McKnight, P. E., & Kashdan, T. B. (2009). The importance of functional impairment to mental health outcomes: A case for assessing our goals in depression treatment research. *Clinical Psychology Review*, *29*, 243-259. doi: 10.1016/j.cpr.2009.01.005
- McKnight, P. E., Monfort, S. S., Kashdan, T. B., Blalock, D. V., & Calton, J. M. (2016, April). Anxiety symptoms and functional impairment: A systematic review of the correlation between the two measures. *Clinical Psychology Review*, *45*, 115-130. doi: 10.1016/j.cpr.2015.10.005
- McNatt, D. B. (2000). Ancient Pygmalion joins contemporary management: A meta-analysis of the result. *Journal of Applied Psychology*, *85*, 314-322. doi: 10.1037/0021-9010.85.2.314
- Memon, A., Meissner, C. A., & Fraser, J. (2010). Cognitive interview. A meta-analytic review and study space analysis of the past 25 years. *Psychology, Public Policy, and Law*, *16*, 340-372. doi: 10.1037/a0020518
- Memon, A., Wark, L., Bull, R., & Köhnken, G. (1997). Isolating the effects of the cognitive interview techniques. *British Journal of Psychology*, *88*, 179-198. doi: 10.1111/j.2044-8295.1997.tb02629.x
- O'Donnell, M. L., Creamer, M., & Pattison, P. (2004). PTSD and depression following trauma: Understanding comorbidity. *American Journal of Psychiatry*, *161*, 1-7. doi: 10.1176/appi.ajp.161.8.1390
- Pedersen, G., & Karterud, S. (2012). The symptom and function dimensions of the Global Assessment of Functioning (GAF) scale. *Comprehensive Psychiatry*, *53*, 292-298. doi: 10.1016/j.comppsy.2011.04.007
- Pereda, N., Guilera, G., Forns, M., & Gómez-Benito, J. (2009). The prevalence of child sexual abuse in community and student samples: A meta-analysis. *Clinical Psychology Review*, *29*, 328-338. doi: 10.1016/j.cpr.2009.02.007
- Pérez-Fuentes, G., Olfson, M., Villegas, L., Morcillo, C., Wang, S., & Blanco, C. (2013). Prevalence and correlates of child sexual abuse: A national study. *Comprehensive Psychiatry*, *54*, 16-27. doi: 10.1016/j.comppsy.2012.05.010
- Raskin, D. C., & Esplin, P. W. (1991). Statement Validity Assessment: Interview procedures and content analysis of children's statements of sexual abuse. *Behavioral Assessment*, *13*, 265-291.
- Rassin, E. (1999). Criteria Based Content Analysis: The less scientific road to truth. *Expert Evidence*, *7*, 265-278. doi: 10.1023/A:1016627527082

- Roma, P., San Martini, P., Sabatello, U., Tatarelli, R., & Ferracuti, S. (2011). Validity of Criteria-Based Content Analysis (CBCA) at trial in free-narrative interviews. *Child Abuse & Neglect*, 35, 613-620. doi: 10.1016/j.chiabu.2011.04.004
- Rosenthal, R. (1984). *Meta-analytic procedures for social research*. Newbury Park, CA: Sage.
- Rosenthal, R. (1994). Parametric measures of effect size. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 231-244). New York, NY: Russell Sage Foundation.
- Sánchez-Meca, J. (1986). La revisión cuantitativa: Una alternativa a las revisiones tradicionales. *Anales de Psicología*, 3, 79-107. Retrieved from <http://www.um.es/metaanalysis/pdf/7004.pdf>
- Sánchez-Meca, J. (2010). Cómo realizar una revisión sistemática y un meta-análisis [How to carry out a systematic review and meta-analysis]. *Aula Abierta*, 38, 53-64. Retrieved from <http://www.um.es/metaanalysis/pdf/5030.pdf>
- Schank, R. C., & Abelson, R. P. (1977). *Scripts, plans, goals and understanding*. Hillsdale, NJ: Lawrence Erlbaum Ass.
- Schelleman-Offermans, K., & Merckelbach, H. (2010). Fantasy proneness as a confounder of verbal lie detection tools. *Journal of Investigative Psychology and Offender Profiling*, 7, 247-260. doi:10.1002/jip.121
- Seijo, A. (2007). *Asignación de credibilidad a los testimonios: Un estudio de las decisiones judiciales*. Trabajo Tutelado de Tercer Ciclo, Universidad de Santiago de Compostela.
- Seijo, M. D., Sousa, M. A., Álvarez, A., Vilariño, M., Abad, M. C., Martínez, M. M.,...Jiménez, M. J. (2014). *Guía para o/a perito psicólogo/a forense*. Retrieved from Colexio Oficial de Psicología de Galicia: <https://copgalicia.gal/guias>
- Smith, N. (2001). *Reading between the lines: An evaluation of the scientific content analysis technique (SCAN)*. Police Research Series paper 135. London: UK Home Office, Research, Development and Statistics Directorate.
- Smith v. State, 259 Ga. 135, 377 S.E.2d 158, 161, cert. denied, 110 S. Ct, 88 (1989).
- Sporer, S. L. (2004). Reality Monitoring and detection of deception. In A. Granhag & L. A. Strömwall (Eds.), *The detection of deception in forensic contexts* (pp. 64-102). Cambridge, UK: Cambridge University Press.
- Steller, M. (1989). Recent developments in statement analysis. In J. C. Yuille (Ed.), *Credibility assessment* (pp. 135-154). Deventer, Holland: Kluwer. doi: 10.1007/978-94-015-7856-1\_8
- Steller, M., & Böhm, C. (2006). Cincuenta años de jurisprudencia del Tribunal Federal Supremo alemán sobre la psicología del testimonio. Balance y perspectiva [Fifty years of the German Federal Court jurisprudence on forensic psychology]. In T. Fabian, C. Böhm, & J. Romero (Eds.), *Nuevos caminos y conceptos en la psicología jurídica* (pp. 53-67). Münster, Germany: LIT Verlag.

- Steller, M., & Boychuk, T. (1992). Children as witness in sexual abuse cases: Investigative interview and assessment techniques. In H. Dent & R. Flin (Eds.), *Children as witnesses* (pp. 47-73). New York, NY: John Wiley & Sons, Inc.
- Steller, M., & Köhnken, G. (1989). Criteria-Based Content Analysis. In D. C. Raskin (Ed.), *Psychological methods in criminal investigation and evidence* (pp. 217-245). New York, NY: Springer-Verlag.
- Stoltenborgh, M., van Ijzendoorn, M. H., Euser, E. M., & Bakermans-Kranenburg, M. J. (2011). A global perspective on child sexual abuse: Meta-analysis of prevalence around the world. *Child Maltreatment, 16*, 79-101. doi: 10.1177/1077559511403920
- Ternes, M. (2009). *Verbal credibility assessment of incarcerated violent offenders' memory reports* (Unpublished doctoral thesis). University of British Columbia, Vancouver.
- Thompson, K. M., Crosby, R. D., Wonderlich, S. A., Mitchell, J. E., Redlin, J., Demuth, G., Smyth, J., & Haseltine, B. (2003). Psychopathology and sexual trauma in childhood and adulthood. *Journal of Traumatic Stress, 16*, 35-38. doi: 10.1023/A:1022007327077
- Tukey, J. W. (1960). A survey of sampling from contaminated distributions. In I. Olkin, J. G. Ghurye, W. Hoeffding, W. G. Madow, & H. Mann (Eds.), *Contributions to probability and statistics* (pp. 448-485). Stanford, CA: Stanford University Press.
- Undeutsch, U. (1989). The development of statement reality analysis. In J. Yuille (Ed.), *Credibility assessment* (pp.101-119). Dordrech, Holland: Kluwer Academic Publishers.
- Van Koppen, P. J. (2007). Misapplication of psychology in court. In D. Carson, B. Milne, F.J. Pakes, K. Shalev, & A. Shawyer (Eds.), *Applying psychology to criminal justice* (pp. 265-282). Chichester, UK: John Wiley & Sons. doi: 10.1002/9780470713068.ch16
- Vázquez- Rojas, C. (2014). Sobre la cientificidad de la prueba científica en el proceso judicial [Discussing about the scientificity of the scientific evidence in Court proceedings]. *Anuario de Psicología Jurídica, 24*, 65-73. doi: 10.1016/j.apj.2014.09.001
- Vilariño, M. (2010). *¿Es posible discriminar declaraciones reales de imaginadas y huella psíquica real de simulada en casos de violencia de género?* (Doctoral thesis, Universidad de Santiago de Compostela, Spain). Retrieved from <http://hdl.handle.net/10347/2831>
- Vilariño, M., Arce, R., & Fariña, F. (2013). Forensic-clinical interview: Reliability and validity for the evaluation of psychological injury. *The European Journal of Psychology Applied to Legal Context, 5*, 1-21. Retrieved from [http://scielo.isciii.es/scielo.php?script=sci\\_arttext&pid=S1889-18612013000100001](http://scielo.isciii.es/scielo.php?script=sci_arttext&pid=S1889-18612013000100001)
- Vilariño, M., Novo, M., & Seijo, D. (2011). Estudio de la eficacia de las categorías de realidad del testimonio del Sistema de Evaluación Global (SEG) en casos de violencia de género [Study of the efficacy of the testimony reality categories of the Global Evaluation System (GES) in violence against women cases]. *Revista Iberoamericana de Psicología y Salud, 2*, 1-26. Retrieved from <http://www.redalyc.org/pdf/2451/245116403001.pdf>

- Volbert, R., & Steller, M. (2014). Is this testimony truthful, fabricated, or based on false memory? Credibility assessment 25 years after Steller and Köhnken (1989). *European Psychologist, 19*, 207 -220. doi: 10.1027/1016-9040/a000200
- Vrij, A. (2005). Criteria-Based Content Analysis: A qualitative review of the first 37 studies. *Psychology, Public Policy, and Law, 11*, 3–41. doi: 10.1037/1076-8971.11.1.3
- Vrij, A. (2008). *Detecting lies and deceit: Pitfalls and opportunities* (2nd ed.). Chichester, UK: John Wiley and Sons.
- Vrij, A., Akehurst, L., Soukara, R., & Bull, R. (2004). Detecting deceit via analyses of verbal and nonverbal behavior in children and adults. *Human Communication Research, 30*, 8-41. doi:10.1111/j.1468-2958.2004.tb00723.x
- Vrij, A., Edward, K., Roberts, K.P., & Bull, R. (2000). Detecting deceit via analysis of verbal and nonverbal behavior. *Journal of Nonverbal Behavior, 24*, 239-263. doi:10.1023/A:1006610329284
- World Health Organization. (WHO, 1999, March 29-31). *Report of the consultation on child abuse prevention*. Geneva, Switzerland: Author. Retrieved from <http://www.who.int/iris/handle/10665/65900#sthash.YgBC2YVe.dpuf>
- World Health Organization. (2000). *Women's mental health: An evidence based review*. Geneva, Switzerland: Author.
- Yuille, J. C., Marxsen, D., & Cooper, B. S. (1999). Training investigative interviewers: Adherence to the spirit, as well as the letter. *International Journal of Law and Psychiatry, 22* (3-4), 323-336. doi: 10.1016/S0160-2527(99)00012-6



## 8. ARTÍCULOS DE INVESTIGACIÓN





## 8.1. UNDEUTSCH HYPOTHESIS AND CRITERIA BASED CONTENT ANALYSIS: A META-ANALYTIC REVIEW

Bárbara G. Amado\*, Ramón Arce\*, & Francisca Fariña\*\* (2015)

\*Department of Psicología Organizacional, Jurídico-Forense y Metodología de las Ciencias de Comportamiento, University of Santiago de Compostela (Spain).

\*\*AIPSE Department, University of Vigo (Spain).

### Abstract

The credibility of a testimony is a crucial component of judicial decision-making. Checklists of testimony credibility criteria are extensively used by forensic psychologists to assess the credibility of a testimony, and in many countries they are admitted as valid scientific evidence in a court of law. These checklists are based on the Undeutsch hypothesis asserting that statements derived from the memory of real-life experiences differ significantly in content and quality from fabricated or fictitious accounts. Notwithstanding, there is considerable controversy regarding the degree to which these checklists comply with the legal standards for scientific evidence to be admitted in a court of law (e.g., *Daubert standards*). In several countries, these checklists are not admitted as valid evidence in court, particularly in view of the inconsistent results reported in the scientific literature. Bearing in mind these issues, a meta-analysis was designed to test the Undeutsch hypothesis using the CBCA Checklist of criteria to discern between memories of self-experienced real-life events and fabricated or fictitious accounts. As the original hypothesis was formulated for populations of children, only quantitative studies with samples of children were considered for this study. In line with the Undeutsch hypothesis, the results showed a significant positive effect size that is generalizable to the total CBCA score,  $\delta = 0.79$ . Moreover, a significant positive effect size was observed in each and all of the credibility criteria. In conclusion, the results corroborated the validity of the Undeutsch hypothesis and the CBCA criteria for discriminating between the memory of real self-experienced events and false or invented accounts. The results are discussed in terms of the implications for forensic practice.

**Key words:** meta-analysis, CBCA, credibility, testimony, sexual abuse, child.

### Hipótesis Undeutsch y Criteria Based Content Analysis: Una revision meta-analítica.

#### Resumen

Con frecuencia, la evaluación de la fiabilidad de un testimonio se lleva a cabo mediante el uso de sistemas categoriales de análisis de contenido. Concretamente, el instrumento más utilizado para determinar la credibilidad del testimonio es el Criteria Based Content Analysis

(CBCA), el cual se sustenta en la hipótesis Undeutsch, que establece que las memorias de un hecho auto-experimentado difieren en contenido y calidad de las memorias fabricadas o imaginadas. Las opiniones y resultados contradictorios encontrados en la literatura científica respecto al cumplimiento de los criterios judiciales (*Daubert standards*) así como el abundante número de trabajos existentes sobre la materia, nos llevó a diseñar un meta-análisis para someter a prueba la hipótesis Undeutsch, a través de la validez de los criterios de realidad del CBCA para discriminar entre la memoria de lo auto-experimentado y lo fabricado. Se tomaron aquellos estudios cuantitativos que incluían muestras de menores, esto es, con edades comprendidas entre los 2 y 18 años. En línea con la hipótesis Undeutsch, los resultados mostraron un tamaño del efecto positivo, significativo y generalizable para la puntuación total del CBCA,  $\delta = 0.79$ . Asimismo, en todos los criterios de realidad se encontró un tamaño del efecto positivo y significativo. En conclusión, los resultados avalan la validez de la hipótesis Undeutsch y de los criterios del CBCA para discriminar entre memorias de hechos auto-experimentados y fabricados. Se discuten las implicaciones de los resultados para la práctica forense.

**Palabras clave:** meta-análisis, CBCA, credibilidad, testimonio, abusos sexuales, menores.

### **Hipótese Undeutsch e Criteria Based Content Analysis: Unha revisión meta-analítica**

#### **Resumo**

Con frecuencia, a avaliación da fiabilidade da testemuña lévase a cabo a partires do uso de sistemas categoriais de análise de contido. Concretamente, o instrumento máis usado para determinar a credibilidade do testemuño é o Criteria Based Content Analysis (CBCA), o cal susténtase na Hipótese Undeutsch, que establece que as memorias de feitos auto-experimentados difiren en contido e calidade das memorias fabricadas ou imaxinadas. As opinións e resultados contradictorios atopados na literatura científica respecto ó cumprimento dos criterios xudiciais (*Daubert standards*) así coma o abundante número de traballo existentes sobre a materia, levounos a deseñar un meta-análise para someter a proba a hipótese Undeutsch. A partir da validez dos criterios de realidade do CBCA para discriminar entre a memoria do auto-experimentado e o fabricado. Tomáronse aqueles estudos cuantitativos que incluían mostras de menores, isto é, con idades comprendidas entre os 2 e 18 anos. En liña coa Hipótese Undeutsch, os resultados amosaron un tamaño do efecto positivo, significativo e xeralizable para a puntuación total do CBCA,  $\delta = 0.79$ . asemade, en todos os criterios de realidade atopouse un tamaño do efecto positivo e significativo. En conclusión, os resultados avalan a validez da hipótese Undeutsch e dos criterios do CBCA para discriminar entre memorias de feitos auto-experimentados e fabricados. Discútense as implicacións dos resultados para a práctica forense.

**Palabras chave:** meta-análise; CBCA; credibilidade; testemuño; abusos sexuais; menores.

## Acknowledgements

This research has been carried out within the framework of research project with the Reference Ref: GPC2014/022, funded by the Xunta de Galicia [Galician Autonomous Government] (Spain). The authors acknowledge and thank to Professor Jesús F. Salgado for his methodological support and the revision of this manuscript.

## Introduction

Hans and Vidmar (1986) estimated that in around 85% of judicial cases, the evidence bearing most weight is the testimony, which underscores that the evaluation of a testimony is crucial for judicial judgement-making. In terms of the application of Information Integration Models to legal judgements (Kaplan, 1982), the reliability and validity of the testimony are the mechanisms underlying the evaluation of a testimony. The validity of a testimony, i.e., the value of a testimony for judgement-making is easily estimated and is to be determined by the rulings of judges and the courts. As for the reliability of a testimony, the courts and scientific studies have tended to estimate it in terms of the credibility of a testimony (Arce, Fariña, & Fraga, 2000), which entails the design of methods for its estimation. Traditionally, judges and the courts have performed this function on the basis of legal criteria, jurisprudence, and their own value judgements. Alternatively, numerous scientific techniques (and pseudoscientific) have been proposed such as non-verbal indicators of deception; paraverbal indicators of deception; physiological indicators (e.g., polygraph tests or functional magnetic resonance imaging); and categorical systems of content analysis. Of these, categorical systems of content analysis are currently the most systemically used technique by the courts. Thus, the courts in countries such as Germany, Sweden, Holland, and several states in the USA admit these categorical systems as scientific evidence (Steller & Böhm, 2006; Vrij, 2008). In Spain, where they are also admitted as legally admissible evidence and extensively used by the courts, an analysis of legal judgements showed that when a forensic psychological report based on a categorical system of content analysis (i.e., Statement Validity Analysis, SVA) confirmed the credibility of a testimony, the conviction rate was 93.3%, but when it failed to do so, the acquittal rate was 100%. In contrast, in other countries such as the UK, the US, and Canada these checklists are not admitted as legally valid evidence (Novo & Seijo, 2010).

Underlying categorical content systems is what is commonly referred to as the *Undeutsch hypothesis* that asserts that the memory of a real-life self-experienced event differs in content and quality from a fabricated or imagined event (Undeutsch, 1967, 1989). On the basis of this hypothesis, Steller and Köhnken (1989) have integrated all the categorical systems (e.g., Arntzen, 1970; Dettenborn, Froehlich, & Szewczyk, 1984; Szewczyk, 1973; Undeutsch, 1967) into what is known as Criteria Based Content Analysis (CBCA), which has become the leading categorical system for evaluating the credibility of a testimony (Griesel, Ternes, Schraml, Cooper, & Yuille, 2013; Vrij, 2008).

CBCA, which is part of SVA, consists of three elements: 1) semi-structured interview, i.e., the free narrative interview; 2) content analysis on CBCA criteria; and 3) evaluation of CBCA outcomes using the Validity Checklist. The semi-structured interview involves a narrative format that, unlike other types of interview such as standard, interrogative or structured interviews, facilitates the emergence of criteria (Vrij, 2005). Moreover, this type of

interview generates more information (Memon, Meissner, & Fraser, 2010), which meets the requirement that CBCA criteria content analysis be performed on sufficient material (Köhnken, 2004; Steller, 1989).

The Checklist of CBCA criteria (Steller & Köhnken, 1989) consists of 19 criteria structured around 5 major categories: general characteristics, specific contents, peculiarities of the content, contents related to motivation, and specific elements of aggression (see Table 1). These criteria of reality do not constitute a methodic categorical system (Bardin, 1977; Weick, 1985), but rather stem from the authors' personal experiences of cases (Steller & Köhnken, 1989). Though this checklist was originally developed as a comprehensive system of credibility criteria grounded on the Undeutsch hypothesis, Raskin, Esplin, and Horowitz (1991) highlighted that only the first 14 criteria are related to the Undeutsch hypothesis, and the remaining 5 criteria are not associated to the aforementioned hypothesis as they are not linked to the concept of memory of actual events. This reclassification overlaps, though not entirely, with the theoretical model proposed by Köhnken (1996) who regroups these major categories into two main factors: cognitive (criteria 1 to 13), and motivational (criteria 14 to 18). The cognitive factor encompasses cognitive and verbal skills, and implies a self-experienced statement contains CBCA criteria from 1 to 13. The motivational factor, however, relies on the individual's ability to avoid appearing deceitful and ways of managing a positive self-impression of oneself as an honest witness. Thus, the motivational factor covers criteria 14 to 18, which are contrary-to-truthfulness-stereotype criteria though they really appear in true statements. Thus, these criteria have been suggested to be useful for assessing the hypothesis of the (partial) fabrication of statements (Köhnken, 1996, 2004).

**Table 1. CBCA-Criteria (adapted from Steller & Köhnken, 1989)**

<p><b>GENERAL CHARACTERISTICS</b></p> <ol style="list-style-type: none"> <li>1. Logical structure</li> <li>2. Unstructured production</li> <li>3. Quantity of details</li> </ol>
<p><b>SPECIFIC CONTENTS</b></p> <ol style="list-style-type: none"> <li>4. Contextual embedding</li> <li>5. Descriptions of interactions</li> <li>6. Reproduction of conversation</li> <li>7. Unexpected complications during the incident</li> </ol>
<p><b>PECULIARITIES OF CONTENT</b></p> <ol style="list-style-type: none"> <li>8. Unusual details</li> <li>9. Superfluous details</li> <li>10. Accurately reported details misunderstood</li> <li>11. Related external associations</li> <li>12. Accounts of subjective mental states</li> <li>13. Attribution of perpetrator's mental state</li> </ol>
<p><b>MOTIVATION-RELATED CONTENTS</b></p> <ol style="list-style-type: none"> <li>14. Spontaneous corrections</li> <li>15. Admitting lack of memory</li> <li>16. Raising doubts about one's own testimony</li> <li>17. Self-deprecation</li> <li>18. Pardoning the perpetrator</li> </ol>
<p><b>OFFENCE-SPECIFIC ELEMENTS</b></p> <ol style="list-style-type: none"> <li>19. Details characteristic of the offence</li> </ol>

Initially, the CBCA criteria were intended for populations of child alleged victims of sexual abuse. However, CBCA criteria have been applied to other types of events and age ranges. This generalization has been extended to professional practice too. Thus, the guidelines of the Institute of Forensic Medicine in Spain, which is the official public institution responsible for forensic evidence, recommends SVA as part of the protocol for women alleging intimate partner violence (Arce & Fariña, 2012). Moreover, there is no consensus regarding the term *minor*, particularly since studies using the term range from 2 to 18-year-olds and the concept of minor is generally associated to the legal age of criminal responsibility. In relation to the context of application, only field studies involve real cases of sexual abuse, since it would be unethical to subject children to conditions or instructions of victims of sexual abuse. Hence, most research is experimental and certain authors have expressed their reservations regarding validity (Konecni & Ebbesen, 1992). Moreover, real eyewitnesses and subjects under high fidelity laboratory conditions have been found to perform different tasks (Fariña, Arce, & Real, 1994). In order to overcome this limitation, some experimental studies have recreated high fidelity simulated conditions in order to mimic the context of recall of child alleged victims of sexual abuse. These conditions have been defined as personal involvement, negative emotional tone of an event, and extensive loss of control over the situation (Steller, 1989). Accordingly, this achieves face validity, with external validity remaining entirely untested (Konecni & Ebbesen, 1992). Nevertheless, in spite of the weaknesses of this experimental paradigm in generalizing CBCA outcomes to the forensic context, experimental laboratory studies are useful for assessing certain variables that may lead to further internal validity research (Griesel et al., 2013). In contrast, the limitation of field studies resides in their difficulty in sustaining the ground truth in accurate objective criteria. These differences in research paradigms imply more credibility criteria are observed in field studies than in experimental ones (Vrij, 2005).

The CBCA criteria are measured on two response scales, presence vs. absence, and the degree of presence. The unit of analysis is the full statement for the first major category, *general characteristics*, and for the remainder frequency counts. The presence of reality criteria is assumed to be indicative of memory based on real-life events, but the absence of criteria does not imply recall is based on fabricated accounts. Additionally, fictitious memory may contain reality criteria. Thus, the evaluation rests on a clinical judgement (Köhnken, 2004), which is semi-objective. Nonetheless, a replicable objective evaluation system, i.e., stringent is a fundamental standard for forensic practice. Alternatively a variety of decision rules have been proposed such as the presence of 3 criteria to judge a statement as true (Arntzen, 1983); of 7 criteria, criteria 1 to 5, plus 2 others; or the first three ones, plus 4 others (Zaparniuk, Yuille, & Taylor, 1995). Unfortunately, these decision rules lack empirical rigour.

Furthermore, the application of criteria derived from the Undeutsch hypothesis to the forensic assessment of the credibility of a testimony (to be more precise, the truthfulness of a testimony given that credibility is a legal concept defined by the ruling of judges and the courts) must fulfil the legal standards for scientific evidence to be admitted as such in court. The standards governing the admission of expert testimony in court have been laid down by the Supreme Court of the United States in *Daubert v. Merrel Dow Pharmaceuticals* (1993) and are as follows: 1) Is the scientific hypothesis testable?; 2) has the proposition been tested?; 3) is there a known error rate?; 4) has the hypothesis and/or technique been subjected to peer review and publication; and 5) is the theory upon which the hypothesis and/or

technique is based on generally accepted among the appropriate scientific community? Several authors have raised their doubts on whether SVA/CBCA fulfil the above criteria, and this has led to seemingly contradictory viewpoints (Honts, 1994; Vrij, 2008). Bearing in mind these contradictory results and interpretations, and the wealth of scientific evidence, the aim of this meta-analysis was to review the literature on both experimental and field studies in order to assess the degree to which the Undeutsch hypothesis meets the judicial standards of evidence by estimating the effect size of the CBCA criteria. As initially the hypothesis was formulated for a population of children, though it has also been applied to adults, this review was restricted to studies on samples of children.

## Method

### Literature search

The aim of the scientific literature search was to identify all of the empirical studies assessing the efficacy of CBCA criteria in discriminating between true statements of actual experiences, and the invented, imagined, fictitious, fabricated or false accounts of children. An exhaustive multi-method search was undertaken in the following international psychology databases of reference: PsycInfo and all of the databases of Web of Science, the Spanish language databases of reference Psycodoc (database of the Official Spanish College of Psychology), the Italian databases (ACPN, Archivio Collettivo Nazionale dei Periodici), the German Psychlinker databases, and the French human, social sciences, and economics Francis databases; the Google Scholar meta-search engine of scientific articles; a manual search in books; crosschecking all the references included in published reviews of articles and manuals; and directly contacting authors to request copies of unavailable studies. The keywords entered in the search engines were: Criteria Based Content Analysis or CBCA (kriteriumbasierte inhaltsanalyse, análisis de contenido basado en criterios, analisi del contenuto basata su criteri), credibility (glaubwürdigkeit, credibilidad, credibilità), content analysis (inhaltsanalyse, análisis de contenido, analisi del contenuto), child sexual abuse (kindermissbrauch, abusos sexuales a menores, abuso sessuale perpetrato su minori), child testimony (kindliche zeugenaussage, testimonio del menor, bambini testimoni). The searches in the databases of non-English speaking countries were undertaken in the corresponding language. In line with the method of successive approximations, all of the keywords in the selected articles were revised in search of other potential descriptors. However, successive searches with these new descriptors failed to produce any further studies for the meta-analysis.

### Inclusion and exclusion criteria

Though the Undeutsch hypothesis was initially formulated for children, the exact age group it encompasses has never been clearly specified. Nevertheless, as it was intended for judicial contexts and victims of sexual abuse, it is understood that it refers to children under the age of consent. First, the literature has taken the legal concept of minor as below the age of criminal responsibility (< 18 years) (Raskin & Esplin, 1991). Likewise, the lower age group was not related to the model, but the hypothesis is supported in that the memory of genuine life experiences differs in content and quality to memory of fictitious accounts, with



the criteria for discriminating between both types of memory being derived from the witness' verbal account. Thus, the child is expected to have the sufficient narrative capability to express these criteria (Köhnken, 2004). Once again, the lower age limit was crosschecked in the studies reviewed, observing the lowest was a 2-year-old child (Buck, Warren, Betman, & Brigham, 2002; Lamers-Winkelmann & Buffing, 1996), whose narrative skills, memory, gaps in memory, and recovery may be insufficient. Notwithstanding, given that the scientific literature has set a minimum age of 2 years and a maximum of 18 years, all of the studies with witnesses between these ages were included. Thus, studies with samples of children or that calculated the effect size for the subsamples of children were included.

Second, delimiting the testimony to sexual abuse would compel studies to focus on this type of victim, as it would be unethical to subject children to memories of feigned victims. Consequently, the studies reviewed can be subdivided into low fidelity experimental studies (i.e., the scenarios neither involved sexual abuse nor was the implication or motivation of the participants controlled), high fidelity experiments (i.e., the scenarios do not involve sexual abuse, but they create an emotionally charged contexts close to the victimization of sexual abuse, and the implication of the participants is controlled), and field studies (i.e., real cases of sexual abuse where the ground truth is based on judicial judgements, the confession of the accused, medical evidence, and polygraph tests). The effects of the context of the research (i.e., field vs. laboratory high fidelity studies) on the results of the quality of an eyewitness' identification have been found to be significant, and even contradictory (Fariña et al., 1994). Thus, it would be plausible to believe that this same bias may also affect the testimony of children alleged victims of sexual abuse. Hence, according to the circumstances, the context of the research was considered as a moderator.

Third, for the criteria to discriminate between memories of real events and fabricated accounts, SVA proposes statements should be obtained using a free narrative interview e.g., step-wise interview, cognitive interview, Memorandum of Good Practices (Köhnken, 2004; Steller, 1989; Undeutsch, 1989), as they facilitate the emergence of criteria (Vrij, 2005). Likewise, compliance with other SVA criteria was strictly observed, i.e., studies noncompliant with any of the basic characteristics of the interview, such as inappropriate prompts or suggestions, were excluded.

Fourth, the studies admitted as legal evidence should be published in scientific peer-reviewed journals (Daubert v. Merrell Dow Pharmaceuticals, 1993). Notwithstanding, the literature has identified as key references studies that have not been published in these journals (i.e., Boychuk, 1991; Esplin, Houed, & Raskin, 1988), and these have been included in the meta-analysis. Thus, according to the circumstances, compliance with the peer-review publication Daubert standard was taken as a moderator.

Fifth, the effect size was calculated from the data obtained and, if required, the authors were contacted to request the effect size or data for computing it as well as to clarify errors or queries regarding the data.

Sixth, studies with samples shared with other studies were excluded (i.e., Hershkowitz, 1999), to avoid empirical redundancy (duplicity in publishing data) – only the original study and the outlier values  $[IQR \pm 1.5]$  were included. An independent analysis and control of outliers was carried out for each meta-analysis.

A total of 20 publications fulfilled the selection criteria, ranging from 5 effect sizes for *self-deprecation* and *details characteristic of the offence* to 17 effect sizes for the criterion quantity of details (see Table 2), and 22 for the total CBCA score.

## Procedure

Having scanned the literature and selected the articles for this meta-analysis, they were coded according to the variables that could function as a moderator. The literature cites the age of the child, the research paradigm (field studies vs. experimental laboratory studies), and the type of design (within- or between-subject), which may mediate the results. Numerous studies have found a correlation between age and total CBCA score, i.e., the older the child, the greater probability of scoring high on this instrument (Anson, Golding, & Gully, 1993; Buck et al., 2002; Craig, Scheibe, Raskin, Kircher, & Dodd, 1999; Horowitz, Lamb, Esplin, Boychuk, Krispin, & Reiter-Lavery, 1997; Lamb, Sternberg, Esplin, Hershkowitz, & Orbach, 1997; Roma, San Martini, Sabatello, Tatarelli, & Ferracuti, 2011). Moreover, if we take into account the child's development, it is probable that to some extent the presence of certain criteria is influenced by the child's age. In relation to the research paradigm, Vrij (2005) found the differences between statements of genuine experiences and invented ones were greater in field studies than in experimental studies. Nonetheless, the results obtained from both field studies and laboratory experiments corroborated the Undeutsch hypothesis. As for the type of design (within- vs. between-subject), criteria have been found to be sensitive to the type of design, whereas total CBCA scores were not (Bensi, Gambetti, Nori, & Giusberti, 2009). Moreover, within-subject designs enhance the value of CBCA criteria for discriminating between memory of truthful and false statements.

These variables were complemented with others obtained from the content analysis of the studies themselves. Through a method of successive approximations, two raters examined the studies to identify potential moderator variables described in them, that later underwent a *Thurstone style* evaluation by 10 judges on the degree of independence and pertinence for the study aim. Having identified the moderators in the coded studies, the non-productive ones were eliminated,  $p \leq .05$  (for a more detailed description of the method see Arce, Velasco, Novo, & Fariña, 2014). Thus, the following moderators were identified as productive by the raters coding the studies: sex of the participants, the type of scale used for rating the presence of each criterion (i.e., present vs. absent; the weighted presence of a criterion), number of raters, training of the raters, type of interview, coding criteria, and the type of material for content analysis (transcript, video or both).

The coding of these variables (see Appendix 1) was undertaken by two independent raters. Ratings were crosschecked and found to agree totally as was expected, given the clarity and mutual exclusion of the variables.

## Data analysis

The calculation of the effect size was homogenized in Cohen's  $d$ , which was computed, when this statistic was not provided by the author of a study, according to the following statistics and availability: the means and standard deviations/standard mean error and, in absence of these, the  $t$ -test value or the associated probability.

When the results were proportioned in Fisher's  $F$  values, the effect size on  $\eta^2$  was transformed to  $d$ , and only if  $F$  was available without the exact probability ( $d$  being obtained from the exact probability),  $F$  was transformed in  $t$ -scores and, from this,  $d$  was computed.

When the results are provided in proportions an effect size  $\delta$  is obtained (Hedges & Olkin, 1985) using the procedure of Kraemer and Andrews (1982), which is equivalent to Cohen's  $d$ , whereas when the results are expressed in 2X2 contingency tables, phi is obtained, and in turn the effect size of  $d$  (Cohen, 1988; Rosenthal, 1994).

The unit of analysis ( $n$ ) was the number of statements, weighting the estimates of the effect sizes by the number of statements instead of the conventional number of participants.

The meta-analysis was performed according to the procedure of Hunter and Schmidt (2004), with a total of 20 Bare-Bones type meta-analysis: one for each of the CBCA criteria, and one for all of the criteria of a statement.

The relationship between both distributions for which the effect sizes were calculated was crucial for interpreting the effect sizes. As for studies with practical utility, Fritz, Morris, and Richler (2012) recommended three statistics: U1, the Binomial Effect Size Display (BESD), and the Probability of Superiority (PS).

### **Reliability criterion**

The reliability of each criterion was assessed in the primary studies using between-rater reliability or between-rater agreement (e.g., intra-class correlation, Maxwell RE, Pearson's correlation, Cohen's kappa, Fleiss' kappa, concordance index, Spearman's rho). Nevertheless, some studies failed to report estimates of reliability or reported several estimates. In the latter case, the reliability index (between-rater agreement indexes were excluded as they do not measure reliability) best fits the data distribution according to the conclusions of Anson et al. (1993), and Horowitz et al. (1997). Moreover, reliability was not systematically and homogeneously informed (i.e., the statistic provided was either on each criteria and the total criteria, only the total, only the criteria or only the range for each criterion). In short, not all of the studies informed of reliability, and the reliability for each criterion, nor were they estimated using the same statistic. Thus, for criteria 17 to 18 no estimates on the reliability of these measures were found. This prompted us to calculate one estimate of reliability for the criteria, and another for the total CBCA score since the reliability for the criteria is different to that for the whole instrument i.e., the total CBCA score (Horowitz et al., 1997). The estimate of reliability for the individual criteria was calculated using the reliability coefficients of each study to obtain an average  $r$  of .79 ( $SEM = .02$ ). To estimate the reliability of the total CBCA score, the Spearman-Brown prediction formula was used with the  $r$  extracted from .98 of the total score.

## **Results**

The results of the effect sizes calculated for each criterion and for the total CBCA score, the total number of statements, the weighted sample effect size ( $d$ ), the standard deviation, the effect size corrected for criterion unreliability ( $\delta$ ), the percent of variance accounted for by artifacts, and confidence and credibility intervals (when neither of the intervals contained

zero, the estimated effect size is deemed to be significant and generalizable, respectively), are shown in Table 2.

Moreover, the results revealed a large positive effect size for the total CBCA score, for both the weighted sample effect size of  $d$ , 0.78, and the effect size corrected for the criterion unreliability ( $\delta$ ), 0.79; significant, 99%  $CI_d$  [0.72, 0.84], and generalizable, 90%  $CI_\delta$  [0.11, 1.47]; that is, with an expected minimum value of 0.11, and a maximum of 1.47 standard deviation. In practical terms, the results reveal that the rate of correct classifications of the total CBCA score was 68.5% for statements based on genuine experiences (true positives), with a failed detection rate (false negatives) of 31.5% (BESD); and 47% ( $U1 = 47$ ) of the areas covered by both populations did not overlap i.e., they were totally independent; with a .712 probability (PS) of obtaining more CBCA criteria in a statement based on true self-experienced accounts.

**Table 2. Results of the Meta-Analyses for Each CBCA Criterion and the Total CBCA Score**

Criterion	$k$	$N$	$d_w$	$SD_d$	$\delta$	$SD_\delta$	%VE	99% $CI_d$	90% $CI_\delta$
1	16	1381	0.47	0.334	0.52	0.272	46	[0.42, 0.52]	[0.17, 0.87]
2	15	1217	0.40	0.465	0.53	0.588	25	[0.35, 0.45]	[-0.15, 1.22]
3	17	1477	0.77	0.575	0.87	0.594	16	[0.64, 0.89]	[0.10, 1.63]
4	15	1341	0.69	0.307	0.78	0.571	17	[0.64, 0.74]	[0.05, 1.51]
5	16	1407	0.44	0.443	0.50	0.434	25	[0.39, 0.49]	[-0.06, 1.06]
6	16	1407	0.53	0.430	0.59	0.415	27	[0.48, 0.58]	[0.06, 1.12]
7	11	1111	0.29	0.196	0.33	0.000	100	[0.24, 0.34]	[.33]
8	16	1437	0.27	0.366	0.31	0.335	34	[0.23, 0.31]	[-0.12, 0.74]
9	13	1199	0.42	0.323	0.47	0.274	44	[0.37, 0.47]	[0.12, 0.82]
10	13	1062	0.31	0.409	0.35	0.385	30	[0.26, 0.36]	[-0.14, 0.84]
11	10	916	0.28	0.360	0.32	0.328	35	[0.22, 0.34]	[-0.10, 0.74]
12	15	1194	0.46	0.437	0.52	0.419	28	[0.41, 0.51]	[-0.01, 1.06]
13	10	1052	0.18	0.278	0.21	0.222	50	[0.13, 0.23]	[-0.08, 0.49]
14	15	1367	0.20	0.400	0.23	0.383	28	[0.15, 0.25]	[-0.27, 0.72]
15	13	1076	0.15	0.358	0.17	0.318	38	[0.10, 0.20]	[-0.24, 0.58]
16	10	809	0.19	0.308	0.22	0.238	53	[0.02, 0.35]	[-0.09, 0.52]
17	5	447	0.16	0.476	0.18	0.480	20	[0.08, 0.24]	[-0.43, 0.80]
18	6	517	0.23	0.374	0.25	0.343	34	[0.16, 0.30]	[-0.18, 0.69]
19	5	318	1.25	0.773	1.40	0.807	14	[1.10, 1.40]	[0.38, 2.44]
Average			0.40		0.46		35		
Total score	18	1122	0.78	0.587	0.79	0.275	20	[0.72, 0.84]	[0.11, 1.47]

*Note.*  $k$  = number of studies;  $N$  = total sample size;  $d_w$  = effect size weighted for sample size;  $SD_d$  = observed standard deviation of  $d$ ;  $\delta$  = effect size corrected for criterion unreliability;  $SD_\delta$  = standard deviation of  $\delta$ ; %VE = variance accounted by for artifactual errors; 99%  $CI_d$  = 99% confidence interval for  $d$ ; 90%  $CI_\delta$  = 90% credibility interval for  $\delta$ .

Though the results are statistically generalizable, the data from experimental and field studies have, as previously mentioned, practical implications in terms of meeting the Daubert standards. Moreover, concerns have been raised regarding the limitations on generalizing the results of experimental studies, and more criteria have been observed in field studies. Thus, a new meta-analysis was performed on the total CBCA score of studies according to the research paradigm (experimental studies vs. field studies). The results show (see Table 3) the effect size corrected for criterion unreliability ( $\delta$ ) was positive, significant, moderate, 0.56, and generalizable to experimental studies; and positive, significant, more than large (2.71,  $p < .01$ ), and generalizable to field studies. The comparison of the effect sizes of the experimental and field studies, revealed the latter were significantly greater,  $t(7.3) = 3.86$ ,  $p < .01$ ,  $d = 2.17$ ,

than those obtained for the former. The subsequent meta-analysis grouping studies according to the degree of experimental fidelity was not undertaken due to the lack of low fidelity studies ( $k = 1$ ). Thus, as expected, the results of high fidelity studies were practically the same as those for experimental studies ( $\delta = 0.58$ ). The results require the error rate, independence of distributions, and the probability of superiority to be estimated for each research paradigm. For experimental studies, the rate of correct classifications of truthful memories was 63.5%, with a failure rate of 36.5%; an independence of distributions of 36.1%, and the probability of obtaining more criteria in a statement based on memory of real-life events of .654. For field studies, the rate of correct classifications was 90.2%, with a margin of error of 9.8%; the distributions were completely independent in 90.4%; and the probability of superiority was .972.

**Table 3. Results of the Meta-Analyses of Moderators**

Moderators	$k$	$N$	$d_w$	$SD_d$	$\delta$	$SD_\delta$	%VE	99% $CI_d$	90% $CI_\delta$
Field studies	8	413	2.40	1.37	2.71	1.48	9	[2.10, 2.70]	[0.82, 4.60]
Restricted Field studies <sup>1</sup>	6	325	2.33	1.01	2.63	1.05	16	[1.99, 2.66]	[1.28, 3.97]
Experimental studies	12	810	0.50	0.19	0.56	0.00	100	[0.33, 0.66]	[0.56]
High fidelity studies <sup>2</sup>	11	730	0.51	0.20	0.58	0.00	100	[0.34, 0.68]	[0.58]
CBCA 14 criteria version <sup>3</sup>	4	301	0.85	.46	0.96	0.43	29	[0.56, 1.13]	[0.40, 1.52]
CBCA 19 criteria version	14	821	0.84	0.83	0.95	0.88	11	[0.67, 1.01]	[-0.18, 2.07]

*Note.*  $k$  = number of studies;  $N$  = total sample size;  $d_w$  = effect size weighted for sample size;  $SD_d$  = observed standard deviation of  $d$ ;  $\delta$  = effect size corrected for criterion unreliability;  $SD_\delta$  = standard deviation of  $\delta$ ; %VE = variance accounted for by artifactual errors; 99%  $CI_d$  = 99% confidence interval for  $d$ ; 90%  $CI_\delta$  = 90% credibility interval for  $\delta$ .

<sup>1</sup>limited to those with estimations of the reliability of the evaluation, matching groups, and/or independence of measures in the grouping factor (i.e., confirmed vs. non-confirmed cases); <sup>2</sup>recreation of conditions facilitating personal involvement, negative emotional tone of the event, and/or extensive loss of control; <sup>3</sup>CBCA criteria related to Undeutsch hypothesis (Raskin et al., 1991).

In addition to the initial conception that all CBCA criteria are grounded on the Undeutsch hypothesis, an alternative contention sustains that only the first 14 criteria are derived from the above mentioned hypothesis. Thus, a new meta-analysis of the total CBCA score was performed in studies that were restricted to these criteria. The results (see Table 3) showed an effect size corrected for criterion unreliability ( $\delta$ ) positive, significant, large, 0.96, and generalizable.

As for the criteria independently, these results (see Table 2) show a significantly positive effect for all of them with values of effect size corrected for criterion unreliability ( $\delta$ ) ranging from 0.17 for the criterion of reality *admitting lack of memory*, to 1.40 for the criterion *details characteristic of the offence*. Moreover, the effect sizes for the criteria *logical structure*, *quantity of details*, *contextual embedding*, *reproduction of conversations*, *unexpected complications during the incident*, *superfluous details*, and *details characteristic of the offence* were generalizable. Of the literature reviewed, the results contrary to the Undeutsch hypothesis were only found in the criteria *unstructured production*, *unusual details* (Granhag, Strömwall, & Landström, 2006) and *self-deprecation* (Steller, Wellershaus, & Wolf, 1988). For criteria without generalizable results, the moderators could not be assessed due to the lack of information regarding potential moderators in the primary studies with insufficient  $N$ s to proceed.

## Discussion

The results of this study have the following implications in terms of the Undeutsch hypothesis and the Checklist of CBCA criteria. First, the results of the total CBCA score supported the validity of the Undeutsch hypothesis in discriminating between truthful statements based on self-experienced events and fictitious accounts. Hence, the presence of these criteria is associated to truthful statement, and a large and generalizable effect size. Furthermore, no study reported the opposite trend i.e., significantly more criteria in fabricated statements. Thus, the hypothesis is valid and generalizable to other conditions (e.g., children of all ages, different contexts of sexual abuse, and research paradigms). Second, all of the CBCA criteria discriminated significantly between the real-life memories of children and fabricated accounts. Consequently, CBCA criteria were validated for 5 major categories i.e., general characteristics; specific contents; peculiarities of content; motivation-related contents; and offence-specific elements. Third, in relation to the magnitude of the effect size, the criteria *quantity of details* and *details characteristic of the offence* discriminated significantly high with a large effect size. In comparison, the effect size was medium for the categories: *logical structure*; *unstructured production*; *contextual embedding*; *description of interactions*; *reproduction of conversations*; and *accounts of subjective mental states*. Finally, the effect size was small for the remaining categories (i.e., *unexpected complications during the incident*; *unusual details*; *superfluous details*; *accurately reported details misunderstood*; *related external associations*; *attribution of perpetrator's mental state*; *spontaneous corrections*; *admitting lack of memory*; *raising doubts about one's own testimony*; *self-deprecation*; *pardoning the perpetrator*). Fourth, the effect size of the major categories ranged from medium to large (from 0.52 to 0.87) for *general characteristics*; from small to large (from 0.33 to 0.78) for *specific contents*; from small to moderate (from 0.21 to 0.52) for *peculiarities of content*; small for *motivation-related contents* (from 0.17 to 0.25); and more than large ( $> 1.20 SD$ ) for *offence-specific elements*. In other words, the biggest effect size was for *general characteristics*, that bear considerable weight on judgement-making concerning the credibility of a testimony (Zaparniuk et al., 1995), and *offence-specific elements* major categories, that is, elements difficult to fabricate. Fifth, components (Köhnken, 1996), in line with the qualitative findings of Vrij (2005), cognitive criteria (criteria 1 to 13) had larger effect sizes than motivational criteria (criteria 14 to 18) that are not generalizable. If the motivational component is useful for assessing (partially) fabricated memories (alternative hypothesis to the truth), the system is of less value for such a task, and what is more, these criteria are subject to moderators so their use is not generalizable for this function. Sixth, the results shed some light on the degree of compliance with the Daubert standard. In relation to the first and second standards – *Is the scientific hypothesis testable?*, *2) has the proposition been testable?* – this meta-analysis answers the question affirmatively, not only in meeting Daubert standard, but also in validating the hypothesis. Moreover, the error rate was quantified (third Daubert standard). As shown above, three error rates were obtained: a general one for all the studies, a specific one for experimental studies, and one for field studies. The results for all the studies ( $k = 18$ ), and the experimental studies ( $k = 12$ ) were similar to previous reports of more than 30%. In contrast, in field studies ( $k = 8$ ) the error rate fell sharply to 10%. Moreover, 97% of honest statements contained more criteria than false statements. Notwithstanding, the data of field studies has been called into question for neither estimating the reliability of the evaluation of raters, nor matching groups, and the lack of safeguards in the independence of measures in the grouping factor. Though the meta-analytical technique is not concerned with these issues, and includes all of the studies

excluding outliers, a new meta-analysis was undertaken restricted to studies with a grouping factor of confirmed cases vs. fabricated/un-confirmed cases, with reliability estimates for each criteria, and matched groups ( $k = 6$ ), exhibiting a positive, significant and generalizable effect size ( $\delta$ ) of 2.63 (see Table 3). The results of this meta-analysis corroborated robustness was similar to that obtained in all the field studies: a correct classification rate of truthful statements (BESD) of 89.8%, an independence between distributions (U1) of 89.6%, and a probability of superiority (PS) of .969 (i.e., 96.9% of truthful statements contained more criteria than fabricated statements). As for the fourth Daubert standard – *has the hypothesis and/or technique been subjected to peer review and publication?* – the answer is self-evident from the studies themselves. Concerning the fifth criterion – *is the theory upon which the hypothesis and/or technique is based on generally accepted in the appropriate scientific community?* – this meta-analysis has no clear response. Honts (1994) is of the opinion that, save minor controversy, the hypothesis is widely accepted by the scientific community. However, Vrij (2008) has pointed out that this information is unknown since the scientific community was not consulted. Seventh, though the results of experimental studies were similar, they cannot be directly generalized to field studies, and at best exhibit face validity (Konecni & Ebbesen, 1992). Therefore, the findings of experimental studies require field studies to validate them, as they are insufficient on their own.

The results of this meta-analysis are subject to several limitations that should be borne in mind when generalizing the results. First, the reliability of the interview was not systematically contrasted and frequently the interviewers received poor or no training. It is well known that the abilities of the interviewer mediate the contents and quality of the interview (Bembibre & Higuera, 2010; Fisher, Geiselman, & Amador, 1989; Steller & Köhnken, 1989). Second, it is unknown (unreported) if the material for content analysis was sufficient for assessing the veracity of statements (Köhnken, 2004; Steller & Köhnken, 1989), an aspect which is crucial for children since their memory is less productive (Memon, Meissner, & Fraser, 2010). Third, the measures of reliability of the coding rely on the raters themselves without contrasting if reliability is generalizable to other independent raters, but this procedure does not guarantee the data is actually reliable. Reliability in content analysis is derived from the measure of within-rater consistency (of the coder her/himself through time), and between-raters and between-contexts (with other raters independent to the study, and other materials) (Weick, 1985). Moreover, for the criteria in which the unit of analysis was not the entire statement, the estimate of reliability was generally calculated using frequency counts without verifying the exact correspondence between the counts (with some exceptions, Boychuk, Vrij). This practice overestimates reliability. Furthermore, reliability for each of the measures was not systematically reported (e.g., sometimes ranges, total reliability, or only the reliability of the major categories were documented), as well as the coder training that it is related to coding accuracy (Akehurst, Bull, Vrij, & Köhnken, 2004). Fourth, the Undeutsch hypothesis, and thus SVA/CBCA, assert the presence of criteria is indicative of true statements, but conversely, the absence of criteria does not imply a false statement (i.e., there are other alternative hypotheses to a false statement), and numerous studies have used the criteria for classifying false statements, when the categories are operative for classifying real statements but not fabricated ones. This bias arises from the design of experimental studies, in which the researcher has under control the memory of fictitious accounts, but in the forensic evaluation of real cases, hypothesis other than deception must be considered such as the evaluatee's unwillingness to cooperate; insufficient memory of events for an analysis to be undertaken on the credibility of a statement; or impaired cognitive capacity (Köhnken, 2004).

In terms of forensic applications, the analysis of the credibility of a testimony is admissible as incriminating evidence, being inoperative in classifying false statements. Thus, in terms of fidelity with the Undeutsch hypothesis and its application to forensic assessment, the results should be in the direction of the classification of true statements. Fifth, the results of some meta-analysis may be subject to a degree of variability given that  $Ns < 400$  do not guarantee the stability of sampling estimates (Hunter & Schmidt, 2004). Nevertheless, as the results of the inter-meta-analysis were consistent, these may affect the statistical data, with expected effects in terms of the Undeutsch hypothesis. Sixth, the reliability of each criterion was an estimate, given the aforementioned reporting problems in the primary studies. Seventh, the results were somewhat biased towards supporting the hypothesis since some studies failed to publish so called *conflicting data* i.e., data with criteria that failed to discriminate significantly between real-life self-experienced events and fabricated accounts were excluded (e.g., Akehurst, Manton, & Quandt, 2011). In any case, none of these studies reported results that contradicted the Undeutsch hypothesis, but rather criteria that failed to discriminate significantly between both types of memory. Nevertheless, most of these limitations are reflected in the increase in the error variance, reducing the estimated effect sizes which means the true effect sizes would have been greater, lending even more support to the Undeutsch hypothesis.

As for the practical implications of this meta-analysis, the findings support the Undeutsch hypothesis and several Daubert standards, but this does not imply the use of Checklist of CBCA criteria can be directly generalized to the context of forensic evaluation. First, the categorical system proposed is not methodic, i.e., it fails to comply with stringent methodic conditions: mutual exclusion, homogeneity, pertinence, objectivity, fidelity and productivity (Bardin, 1997). For instances, as non-mutual exclusion between categories is guaranteed, the duplicity of measures may arise; the criteria are neither objective nor exhaustive – e.g., Roma et al. (2011) and Horowitz et al. (1997) have proposed the integration or redefinition of criteria due to rating difficulties –; the checklist may need additional criteria; or the checklist lacks internal consistency, in other words, it is not reliable. Second, the forensic application of a checklist of categories is driven by clinical judgements (Köhnken, 2004), or on quantitative decision rules that are not supported by empirical data (Arntzen, 1983; Zaparniuk et al., 1995). However, in the field of forensics an objective and strict decision criterion based on stringent standards of evidence should prevail over subjective clinical judgements, i.e., the rate of classification of false statements as true (false positives) should be 0 (i.e., the burden of proof is on the prosecution; it is entirely inadmissible to present incriminating expert forensic testimony on the basis of unsubstantiated evidences). It would be a decision rule based on data for controlling false positives, and to ensure reliable coding (i.e., within-raters, between-raters, and between-context consistency), which would offset the potential effects of a truth bias or a response bias associated to the application of reality criteria (Griesel et al., 2013; Rassin, 1999; Sporer, 2004). The results of previous meta-analyses have shown it is possible, i.e., in field studies approximately 97% of truthful statements contained more reality criteria than fabricated accounts, with an approximately 90% total independence between the distributions of both groups of statements.



## References

(References marked with an asterisk indicate studies included in the meta-analysis)

- \*Akehurst, L., Bull, R., Vrij, A., & Köhnken, G. (2004). The effects of training professional groups and lay persons to use Criteria-Based Content Analysis to detect deception. *Applied Cognitive Psychology, 18*, 877–891. doi: 10.1002/acp.1057
- \*Akehurst, L., Manton, S., & Quandt, S. (2011). Careful calculation or a leap of faith? A field study of the translation of CBCA ratings to final credibility judgements. *Applied Cognitive Psychology, 25*, 236–243. doi: 10.1002/acp.1669
- Anson, D. A., Golding, S. L., & Gully, K. J. (1993). Child sexual abuse allegations: Reliability of Criteria-Based Content Analysis. *Law and Human Behavior, 17*, 331–341. doi: 10.1007/BF01044512
- Arce, R., & Fariña, F. (2012). Psicología social aplicada al ámbito jurídico [Applied social psychology to the legal context]. In A. V. Arias, J. F. Morales, E. Nouvilas, & J. L. Martínez (Eds.), *Psicología social aplicada* (pp. 157–182). Madrid, Spain: Panamericana.
- Arce, R., Fariña, F., & Fraga, A. (2000). Género y formación de juicios en un caso de violación [Gender and juror judgment making in a case of rape]. *Psicothema, 12*, 623–628.
- Arce, R., Velasco, J., Novo, M., & Fariña, F. (2014). Elaboración y validación de una escala para la evaluación del acoso escolar [Development and validation of a scale to assess bullying]. *Revista Iberoamericana de Psicología y Salud, 5*, 71–104.
- Arntzen, F. (1970). *Psychologie der zeugenaussage. Einführung in die forensische aussagepsychologie* [Psychology of eyewitness testimony. Introduction to forensic psychology of statement analysis]. Göttingen, Germany: Hogrefe.
- Arntzen, F. (1983). *Psychologie der zeugenaussage: Systematik der glaubwürdigkeitsmerkmale* [Psychology of witness statements: The system of reality criteria]. Munich, Germany: C. H. Beck.
- Bardin, L. (1977). *L'Analyse de contenu* [Content analysis]. Paris, France: Presses Universitaires de France.
- Bembibre, J., & Higuera, L. (2010). Eficacia diferencial de la entrevista cognitiva en función de la profesión del entrevistador: Policías frente a psicólogos [Differential efficacy of the cognitive interview as a function of the interviewers' profession: Police vs. psychologists]. In F. Expósito, M. C. Herrera, G. Buena-Casal, M. Novo, & F. Fariña (Eds.), *Psicología jurídica. Áreas de intervención* (pp. 141–149). Santiago de Compostela, Spain: Consellería de Presidencia, Xustiza e Administracións Públicas.
- Bensi, L., Gambetti, E., Nori, R., & Giusberti, F. (2009). Discerning truth from deception: The sincere witness profile. *European Journal of Psychology Applied to Legal Context, 1*, 101–121.
- \*Blandón-Gitlin, I., Pezdek, K., Rogers, M., & Brodie, L. (2005). Detecting deception in children: An experimental study of the effect of event familiarity on CBCA ratings. *Law and Human Behavior, 29*, 187–197. doi: 10.1007/s10979-005-2417-8
- \*Boychuk, T. D. (1991). *Criteria Based Content Analysis of children's statements about sexual abuse: A field-based validation study*. Unpublished doctoral dissertation, Arizona State University.
- Buck, J. A., Warren, A. R., Betman, S. I., & Brigham, J. C. (2002). Age differences in Criteria-Based Content Analysis scores in typical child sexual abuse interviews. *Journal*

- of Applied Developmental Psychology*, 23, 267–283. doi: 10.1016/S0193-3973(02)00107-7
- \*Casado del Pozo, A. M., Romera, R. M., Vázquez, B., Vecina, M., & De Paúl, P. (2004). Análisis estadístico de una muestra de 100 casos de abuso sexual infantil [Statistical analysis of a one hundred real cases of child sexual abuse]. In B. Vázquez (Ed.), *Abuso sexual infantil. Evaluación de la credibilidad del testimonio* (pp. 73–105). Valencia, Spain: Centro Reina Sofía para el Estudio de la Violencia.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: LEA.
- \*Craig, R. A., Scheibe, R., Raskin, D. C., Kircher, J. C., & Dodd, D. H. (1999). Interviewer questions and content analysis of children's statements of sexual abuse. *Applied Developmental Science*, 3, 77–85. doi: 10.1207/s1532480xads0302\_2
- Daubert v. Merrell Dow Pharmaceuticals, 509 U.S. 579 (1993).
- Dettenborn, H., Froehlich, H., & Szewczyk, H. (1984). *Forensische psychologie* [Forensic Psychology]. Berlin, Germany: Deutscher Verlag der Wissenschaften.
- \*Di Blasio, P., & Conti, A. (2000). L'applicazione del "Criteria-Based Content Analysis" (C.B.C.A.) a racconti di storie vere e inventate [Application of Criteria-Based Content Analysis to accounts of real and fabricated stories]. *Maltrattamento e Abuso all'infanzia*, 2(3), 57-78. doi: 10.1400/62968
- \*Erdmann, K., Volbert, R., & Böhm, C. (2004). Children report suggested events even when interviewed in a non-suggestive manner: What are its implications for credibility assessment? *Applied Cognitive Psychology*, 18, 589–611. doi: 10.1002/acp.1012
- \*Esplin, P. W., Houed, T., & Raskin, D. C. (1988). *Application of statement validity assessment*. Paper presented at the NATO Advanced Study Institute on Credibility Assessment, Maratea, Italy.
- Fariña, F., Arce, R., & Real, S. (1994). Ruedas de identificación: De la simulación y la realidad [Linepus: A comparison of high-fidelity research and research in a real context]. *Psicothema*, 6, 395–402.
- Fisher, R. P., Geiselman, R. E., & Amador, M. (1989). Field test of the cognitive interview: Enhancing the recollection of actual victims and witness of crime. *Journal of Applied Psychology*, 74, 722–727. doi: 10.1037/0021-9010.74.5.722
- Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: Current use, calculations, and interpretation. *Journal of Experimental Psychology: General*, 141, 2–18. doi: 10.1037/a0024338
- \*Granhag, P. A., Strömwall, L. A., & Landström, S. (2006). Children recalling an event repeatedly: Effects on RM and CBCA scores. *Legal and Criminological Psychology*, 11, 81–98. doi: 10.1348/135532505X49620
- Griesel, D., Ternes, M., Schraml, D., Cooper, B. S., & Yuille, J. C. (2013). The ABC's of CBCA: Verbal credibility assessment in practice. In B. S., Cooper, D. Griesel, & M. Ternes (Eds.), *Applied issues in investigative interviewing, eyewitness memory, and credibility assessment* (pp. 293–323). New York, NY: Springer. doi: 10.1007/978-1-4614-5547-9\_12
- Hans, V. P., & Vidmar, N. (1986). *Judging the jury*. New York, NY: Plenum Press.
- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. San Diego, CA: Academic Press.
- Hershkowitz, I. (1999). The dynamics of interviews involving plausible and implausible allegations of child sexual abuse. *Applied Developmental Science*, 3, 86–91. doi: 10.1207/s1532480xads0302\_3

- Honts, C. R. (1994). Assessing children's credibility: Scientific and legal issues in 1994. *North Dakota Law Review*, 70, 879–903.
- Horowitz, S. W., Lamb, M. E., Esplin, P. W., Boychuk, T. D., Krispin, O., & Reiter-Lavery, L. (1997). Reliability of criteria-based content analysis of child witness statements. *Legal and Criminological Psychology*, 2, 11–21. doi: 10.1111/j.2044-8333.1997.tb00329.x
- Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting errors and bias in research findings*. Thousand Oaks, CA: Sage.
- Kaplan, M. F. (1982). Cognitive processes in the individual juror. In N. L. Kerr & R. M. Bray (Eds.), *The psychology of the courtroom* (pp. 197–220). New York, NY: Academic Press.
- Köhnken, G. (1996). Social psychology and the law. In G. R. Semin & K. Fiedler (Eds.), *Applied social psychology* (pp. 257–282). Thousand Oaks, CA: Sage.
- Köhnken, G. (2004). Statement Validity Analysis and the detection of the truth. In P. A. Granhag & L. A. Strömwall (Eds.), *The detection of deception in forensic contexts* (pp. 41–63). Cambridge, UK: Cambridge University Press. doi: 10.1017/CBO9780511490071.003
- Köhnken, G., & Steller, M. (1988). The evaluation of the credibility of child witness statements in the German procedural System. *Issues in Legal and Criminological Psychology*, 13, 37–45.
- Konecni, V. J., & Ebbesen, E. B. (1992). Methodological issues on legal decision-making, with special reference to experimental simulations. In F. Lösel, D. Bender, & T. Bliesener (Eds.), *Psychology and law. International perspectives* (pp. 413–423). Berlin, Germany: Walter de Gruyter.
- Kraemer, H. C., & Andrews, G. (1982). A non-parametric technique for meta-analysis effect size calculation. *Psychological Bulletin*, 91, 404–412. doi: 10.1037/0033-2909.91.2.404
- \*Lamb, M. E., Sternberg, K. J., Esplin, P. W., Hershkowitz, I., & Orbach, Y. (1997). Assessing the credibility of children's allegations of sexual abuse: A survey of recent research. *Learning and Individual Differences*, 9, 175–194. doi: 10.1016/S1041-6080(97)90005-4
- Lamers-Winkelmann, F., & Buffing, F. (1996). Children's testimony in the Netherlands: A study of Statement Validity Analysis. *Criminal Justice and Behavior*, 23, 304–321. doi: 10.1177/0093854896023002004
- \*Mazzoni, G., & Ambrosio, K. (2002). L'analisi del resoconto testimoniale in bambini: Impiego del metodo di analisi del contenuto C.B.C.A. in bambini di 7 anni [Assessment of child witness statements: Application of CBCA method in a 7 year old children sample]. *Psicologia e Giustizia*, 3 (2).
- \*Mazzoni, G., & Pezzati, S. (2002). Esame della validità del C.B.C.A. in racconti di bambini di 4-5 anni [An exam of validity of the CBCA in the account's four-five year old children.]. *Eta' Evolutiva*, 73, 5–17.
- Memon, A., Meissner, C. A., & Fraser, J. (2010). Cognitive interview. A meta-analytic review and study space analysis of the past 25 years. *Psychology, Public Policy, and Law*, 16, 340–372. doi: 10.1037/a0020518
- Novo, M., & Seijo, D. (2010). Judicial judgement-making and legal criteria of testimonial credibility. *European Journal of Psychology Applied to Legal Context*, 2, 91–115.
- Raskin, D. C., & Esplin, P. W. (1991). Statement Validity Assessment: Interview procedures and content analysis of children's statements of sexual abuse. *Behavioral Assessment*, 13, 265–291.

- Raskin, D. C., Esplin, F. W., & Horowitz, S. (1991). *Investigative interviews and assessment of children in sexual abuse cases*. Unpublished manuscript, Department of Psychology, University of Utah.
- Rassin, E. (1999). Criteria Based Content Analysis: The less scientific road to truth. *Expert Evidence*, 7, 265–278. doi: 10.1023/A:1016627527082
- \*Roma, P., San Martini, P., Sabatello, U., Tatarelli, R., & Ferracuti, S. (2011). Validity of Criteria-Based Content Analysis (CBCA) at trial in free-narrative interviews. *Child Abuse & Neglect*, 35, 613–620. doi: 10.1016/j.chiabu.2011.04.004
- Rosenthal, R. (1994). Parametric measures of effect size. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 231–244). New York, NY: Russell Sage Foundation.
- \*Santtila, P., Roppola, H., Runtti, M., & Niemi, P. (2000). Assessment of child witness statements using Criteria-Based Content Analysis (CBCA): The effects of age, verbal ability, and interviewer's emotional style. *Psychology, Crime & Law*, 6, 159–179. doi: 10.1080/10683160008409802
- Sporer, S. L. (2004). Reality monitoring and detection of deception. In P. A. Granhag & L. A. Strömwall (Eds.), *The detection of deception in forensic contexts* (pp. 64–102). Cambridge, UK: Cambridge University Press. doi: 10.1017/CBO9780511490071.004
- Steller, M. (1989). Recent developments in statement analysis. In J. C. Yuille (Ed.), *Credibility assessment* (pp. 135–154). Deventer, Holland: Kluwer. doi: 10.1007/978-94-015-7856-1\_8
- Steller, M., & Böhm, C. (2006). Cincuenta años de jurisprudencia del Tribunal Federal Supremo alemán sobre la psicología del testimonio. Balance y perspectiva [Fifty years of the Federal German Court jurisprudence about forensic psychology]. In T. Fabian, C. Böhm, & J. Romero (Eds.), *Nuevos caminos y conceptos en la psicología jurídica* (pp. 53–67). Münster, Germany: LIT Verlag.
- Steller, M., & Köhnken, G. (1989). Criteria-Based Content Analysis. In D. C. Raskin (Ed.), *Psychological methods in criminal investigation and evidence* (pp. 217–245). New York, NY: Springer-Verlag.
- \*Steller, M., Wellershaus, P., & Wolf, T. (1988). *Empirical validation of Criteria-Based Content Analysis*. Paper presented at the NATO Advanced Study Institute on Credibility Assessment, Maratea, Italy.
- \*Strömwall, L. A., Bengtsson, L., Leander, L., & Granhag, P. A. (2004). Assessing children's statements: The impact of a repeated experience on CBCA and RM ratings. *Applied Cognitive Psychology*, 18, 653–668. doi: 10.1002/acp.1021
- Szewczyk, H. (1973). Kriterien der Beurteilung kindlicher zeugenaussagen [Criteria for the evaluation of child witnesses]. *Probleme und Ergebnisse der Psychologie*, 46, 47–66.
- \*Tye, M. C., Amato, S. L., Honts, C. R., Devitt, M. K., & Peters, D. (1999). The willingness of children to lie and the assessment of credibility in an ecologically relevant laboratory setting. *Applied Developmental Science*, 3, 92–109. doi: 10.1207/s1532480xads0302\_4
- Undeutsch, U. (1967). Beurteilung der glaubhaftigkeit von aussagen [Evaluation of statement credibility/ Statement validity assessment]. In U. Undeutsch (Ed.), *Handbuch der Psychologie*, Vol. 11: Forensische Psychologie (pp. 26–181). Göttingen, Germany: Hogrefe.
- Undeutsch, U. (1989). The development of statement reality analysis. In J. Yuille (Ed.), *Credibility assessment* (pp.101–119). Dordrech, Holland: Kluwer Academic Publishers.
- Vrij, A. (2005). Criteria-Based Content Analysis: A qualitative review of the first 37 studies. *Psychology, Public Policy, and Law*, 11, 3–41. doi: 10.1037/1076-8971.11.1.3

- Vrij, A. (2008). *Detecting lies and deceit: Pitfalls and opportunities* (2nd ed.). Chichester, UK: John Wiley and Sons.
- \*Vrij, A., Akehurst, L., Soukara, S., & Bull, R. (2002). Will the truth come out? The effect of deception, age, status, coaching and social skills on CBCA scores. *Law and Human Behavior*, *26*, 261–283. doi: 10.1023/A:1015313120905
- \*Vrij, A., Akehurst, L., Soukara, R., & Bull, R. (2004). Detecting deceit via analyses of verbal and nonverbal behavior in children and adults. *Human Communication Research*, *30*, 8–41. doi: 10.1111/j.1468-2958.2004.tb00723.x
- Weick, K. E. (1985). Systematic observational methods. In G. Lindzey & E. Aronson (Eds.), *The handbook of social psychology* (Vol. 1, pp. 567–634). Hillsdale, NJ: LEA.
- Zaparniuk, J., Yuille, J. C., & Taylor, S. (1995). Assessing the credibility of true and false statements. *International Journal of Law and Psychiatry*, *18*, 343–352. doi: 10.1016/0160-2527(95)00016-B



Appendix 1. Moderator Variables

Primary studies	N	Age	Sex	Paradigm	Design	Raters	Coding Scale	Coders training	Type of interview	Criteria	Transcript/ video
Akehurst, Bull, Vrij, and Köhnken (2004)	151	7-11	M: 23 F: 26	Experimental: active	Within	58	0-4	Extensive training in CBCA coding	Step-wise interview	13	Transcript
Akehurst, Bull, Vrij, and Köhnken (2004)	132			Experimental: video	Within					13	Transcript
Akehurst, Manton, and Quandt (2011) <sup>2,6</sup>	31	6-17	M: 5 F: 26	Field	Between	2	1-5	Extensive training in CBCA coding	Step-wise interview	3	Transcript
Blandon-Gitlin, Pezdek, Rogers, and Brodie (2005) <sup>5</sup>	94	9-12	-	Experimental: active	Between	2	0-1	Extensive training in CBCA coding	Step-wise interview	16	Transcript
Boychuk (1991)	75	4-16	M: 15 F: 60	Field	Between	2	0-1	Expert raters	No standardized interview procedures	19	Transcript
Casado del Pozo, Romera, Vázquez Mezquita, Vecina, and de Paúl (2002) <sup>6</sup>	96	4-18	M: 28 F: 72	Field	Between	2	0-1	Expert raters	SVA guidelines	19	-
Craig, R. A., Scheibe, R., Raskin, D. C., Kircher, J. C., y Dodd, D. H. (1999) <sup>1</sup>	48	3-16	M: 11 F: 37	Field	Between	4	0-1	8 hours training	SVA guidelines	14	Transcript
Di Blasio and Conti (2000)	88	9	M: 25 F: 19	Experimental: memory	Within	2	0-1	Expert raters	Step-wise interview	19	-
Erdmann, Volbert, and Böhm (2004)	70	6-8	M: 36 F: 31	Experimental: memory	Within	2	0-1/0-2	Experts raters	Step-wise interview	15	Transcript
Esplin, Houed, and Raskin (1988)	40	3-15	-	Field	Between	1	0-2	Intensive training in CBCA coding	-	19	Transcript
Granhag, Strömwall, and Landström (2006)	80	12-13	M: 42 F: 38	Experimental: staged	Between	2	0-2	Intensive training in CBCA coding	Cognitive Interview	10	-
Lamb, Sternberg, Esplin, Orbach, and Hovav (1997) <sup>6</sup>	89	4-13	M: 28 F: 70	Field	Between	2	0-1	Intensive training in CBCA coding	No standardized interview procedures.	14 <sup>1</sup>	Transcript
<b>Primary studies</b>	<b>N</b>	<b>Age</b>	<b>Sex</b>	<b>Paradigm</b>	<b>Design</b>	<b>Raters</b>	<b>Coding Scale</b>	<b>Coders training</b>	<b>Type of interview</b>	<b>Criteria</b>	<b>Transcript/ video</b>
Mazzoni and Pezzati (2002) <sup>5</sup>	60	4-5	M: 20 F: 21	Experimental: memory	Within	2	0-2	Training in CBCA	Step-wise interview	19	Transcript

Mazzoni and Ambrosio (2002) <sup>5</sup>	60	7	-	Experimental: memory	Within	2	0-2	-	Step-wise interview	19	Transcript
Roma, San Martini, Sabatello, Tatarelli, and Ferracuti (2011) <sup>1,6</sup>	109	4-14	M: 23 F: 86	Field	Between	2	0-1	Expert raters	Step-wise interview	14	Transcript
Santtila, Roppola, Runtti, and Niemi (2000) <sup>1,5</sup>	136	7-14	M: 34 F: 34	Experimental: memory	Within	2	0-1/0-2	Expert raters	Step-wise interview	14	Transcript
Steller, Wellershaus, and Wolf (1988)	176	10-13	-	Experimental: memory	Within	3	0-3	-	SVA guidelines	16	-
Strömwall, Bengtsson, Leander, and Granhag (2004) <sup>3,5</sup>	41	10-13	M: 45 F: 42	Experimental: active	Between	2	0-1/0-2	Extensive training in CBCA coding	Cognitive interview	15	Transcript
Strömwall, Bengtsson, Leander, and Granhag (2004) <sup>4,5</sup>	46										
Tye, Amato, Honts, Devitt, and Peters (1999)	28	6-10	M: 21 F: 27	Experimental: active	Between	3	0-2	Expert raters	SVA guidelines	12	Transcript
	36	5-6	M: 16 F: 20								
Vrij, Akehurst, Soukara, and Bull (2002) <sup>5</sup>	56	10-11	M: 22 F: 34	Experimental: active	Between	2	1-5	Extensive training in CBCA coding	Step-wise interview	9	Transcript
	57	14-15	M: 33 F: 24								
	35	5-6	M: 16 F: 19								
Vrij, Akehurst, Soukara, and Bull (2004) <sup>5</sup>	54	10-11	M: 22 F: 32	Experimental: active	Between	2	1-5	Extensive training in CBCA coding	Step-wise interview	9	Transcript
	55	14-15	M: 32 F: 23								

Note. <sup>1</sup>CBCA14 criteria version; <sup>4</sup>limited to those criteria discriminating significantly between self-experienced and fabricated statements; <sup>3</sup>event experienced once, <sup>4</sup>event experienced four times; <sup>5</sup>high fidelity study, <sup>6</sup>restricted field study





## 8.2. CRITERIA-BASED CONTENT ANALYSIS (CBCA) REALITY CRITERIA IN ADULTS: A META-ANALYTIC REVIEW

Bárbara G. Amado\*, Ramón Arce\*, Francisca Fariña\*\*, & Manuel Vilariño\* (2016)

\* University of Santiago de Compostela (Spain).

\*\* University of Vigo (Spain).

### Abstract

Criteria-Based Content Analysis (CBCA) is the tool most extensively used worldwide for evaluating the veracity of a testimony. CBCA, initially designed for evaluating the testimonies of victims of child sexual abuse, has been empirically validated. Moreover, CBCA has been generalized to adult populations and other contexts though this generalization has not been endorsed by the scientific literature. Thus, a meta-analysis was performed to assess the Undeutsch hypothesis and the CBCA checklist of criteria in discerning in adults between memories of self-experienced real-life events and fabricated or fictitious memories. Though the results corroborated the Undeutsch Hypothesis, and CBCA as a valid technique, the results were not generalizable, and the *self-deprecation* and *pardoning the perpetrator* criteria failed to discriminate between both memories. The technique can be complemented with additional reality criteria. The study of moderators revealed discriminating efficacy was significantly higher in field studies on sexual offences and intimate partner violence. The findings are discussed in terms of their implications as well as the limitations and conditions for applying these results to forensic settings.

**Keywords:** Criteria-Based Content Analysis; adults; statements; credibility; meta-analysis.

### Criterios de realidad del CBCA en adultos: Una revisión meta-analítica

#### Resumen

El Criteria-Based Content Analysis (CBCA) constituye la herramienta mundialmente más utilizada para la evaluación de la credibilidad del testimonio. Originalmente fue creado para testimonios de menores víctimas de abuso sexual, gozando de amparo científico. Sin embargo, se ha generalizado su práctica a poblaciones de adultos y otros contextos sin un aval de la literatura para tal generalización. Por ello, nos planteamos una revisión meta-analítica con el objetivo de contrastar la Hipótesis Undeutsch y los criterios de realidad del CBCA para conocer su potencial capacidad discriminativa entre memorias de eventos auto-experimentados y fabricados en adultos. Los resultados confirman la hipótesis Undeutsch y validan el CBCA como técnica. No obstante, los resultados no son generalizables y los criterios *autodesaprobación* y *perdón al autor del delito* no discriminan entre ambas memorias. Además, se encontró que la técnica puede ser complementada con criterios adicionales de realidad. El estudio de moderadores mostró que la eficacia discriminativa era significativamente superior en estudios de campo en casos de violencia sexual y de género. Se

discute la utilidad, así como las limitaciones y condiciones para la transferencia de estos resultados a la práctica forense.

**Palabras clave:** Criteria-Based Content Analysis; adultos; declaraciones; credibilidad; meta-análisis.

## **Crterios de realidade do CBCA en adultos: Unha revisión meta-analítica**

### **Resumo**

O Criteria-Based Content Analysis (CBCA) constitúe a ferramenta mundialmente máis utilizada para a avaliación da credibilidade do testemuño. Orixinariamente foi creado para testemuños de menores vítimas de abuso sexual, gozando de respaldo científico. Nembargantes, xeralizouse a súa práctica a poboacións de adultos e outros contextos sen un aval da literatura para tal xeralización. Por isto, nos plantexámos unha revisión meta-analítica co obxecto de contrastar a Hipótese Undeutsch e os criterios de realidade do CBCA para coñecer a súa potencial capacidade discriminativa entre memorias de feitos auto-experimentados e fabricados en adultos. Os resultados confirman a hipótese Undeutsch e validan o CBCA como técnica. Non obstante, os resultados non son xeralizables e os criterios auto-desaprobación e perdón ó autor do delito non discriminan entre ambas memorias. Con todo, achouse que a técnica pode ser complementada con criterios adicionais de realidade. O estudo de moderadores mostrou que a eficacia discriminativa era significativamente superior nos estudos de campo en casos de violencia sexual e de xénero. Discútese a súa utilidade, así como as limitacións e condicións para a transferencia destes resultados á práctica forense.

**Palabras chave:** Criteria-Based Content Analysis; adultos; declaracións; credibilidade; meta-análise.

### **Introduction**

The credibility of a testimony, primarily the victim's and in particular in relation to crimes committed in private (e.g., sexual offenses, domestic violence), is the key element determining legal judgements (Novo & Seijo, 2010), affecting an estimated 85% of cases worldwide (Hans & Vidmar, 1986). Though an array of tools for evaluating credibility have been designed and tested (Vrij, 2008), Criteria-Based Content Analysis [CBCA] (Steller & Köhnken, 1989) remains the technique of choice, enjoys wide acceptance among the scientific community (Amado, Arce, & Fariña, 2015), and is admissible as valid evidence in the law courts of in several countries (Steller & Böhm, 2006; Vrij, 2008). Though the technique was initially designed to be applied to the testimony of victims of child abuse sexual, its application has been extended to adults, witnesses, offenders, and other case types by Forensic Psychology Institutes in judicial proceedings (Arce & Fariña, 2012). The meta-analysis of Amado et al. (2015) found that the technique underpinning the Undeutsch

Hypothesis (1967) that contends that memories of self-experienced events differ in content and quality to memories of fabricated or fictitious accounts, was equally valid in other contexts and age ranges up to the age of 18 years. Prior to the present review, empirical studies had already contrasted the validity of the Hypothesis in adult populations and in different contexts (Vrij, 2005, 2008). Moreover, as the Hypothesis was grounded on memory content, it had been theoretically advanced that the Hypothesis would be equally applicable to adults and contexts different to sexual abuse (Berliner & Conte, 1993).

CBCA consists of 19 reality criteria which are grouped into two factors: cognitive (criteria 1 to 13), and motivational (criteria 14 to 18). According to the original formulation, both factors are underpinned by the Undeutsch Hypothesis, but Raskin, Esplin and Horowitz (1991) have underscored that only 14 conform to the aforementioned Hypothesis (14-criteria version).

**Table 1. Additional criteria**

- |   |
|---|
| <ul style="list-style-type: none"> <li>• Reporting style (is long-winded when interviewee described irrelevant aspects that were not asked).</li> <li>• Display insecurities (uncertainty about the description of an item).</li> <li>• Providing reasons lack memory (express reasons for not being able to give a detailed description).</li> <li>• Clichés (expressions or utterances that introduce delays into the report).</li> <li>• Repetitions (elements already described were repeated without additional details).</li> </ul> |
|---|

CBCA has encompassed additional categories, some applicable to all contexts (Table 1) (Höfer, Köhnken, Hanewinkel, & Bruhn, 1993), and others for specific cases (Arce & Fariña, 2009; Juárez, Mateu, & Sala, 2007; Volbert & Steller, 2014), which may be combined with other techniques with diverse theoretical underpinnings such as memory attributes (Vrij, 2008).

CBCA is extensively used in forensic practice as a tool for discriminating the memories of adults of self-experienced and fabricated events in different case types. However, due to the numerous inconsistencies in the literature (e.g., designs failing to meet the requirements for applying CBCA, conclusions of non-significant effects not substantiated by the data given the poor statistical power of the studies,  $1-\beta < .80$ ), and the contradictory use of CBCA in adults, a meta-analysis was performed to assess the Undeutsch Hypothesis in an adult population; the discriminating efficacy of CBCA and additional reality criteria; and the effect of the context (case type), lie coaching effect, witness status, and the research paradigm.

## Method

### Literature search

An extensive scientific literature search was undertaken to identify empirical studies applying content analysis to adult testimony in order to discriminate between self-experienced and fabricated statements, be they deliberately invented or implanted memories. The literature search consisted of a multimethod approach to meta-search engines (Google, Google Scholar, Yahoo); world leading scientific databases (PsycInfo, MedLine, Web of Science, Dissertation

Abstracts International); academic social networks for the exchange of knowledge in the scientific community (i.e., Researchgate, Academia.edu); *ancestry approach* (crosschecking the bibliography of the selected studies); and contacting researchers to request unpublished studies mentioned in published studies. A list of descriptors was generated for successive approximations (i.e., the descriptors of the keywords in the selected articles were included): reality criteria, content analysis, verbal cues, verbal indicators, testimony, CBCA, Criteria Based Content Analysis, credibility, adult, statement, allegation, deception, detection, lie detection, truthful account, Statement Validity Assessment, SVA. These descriptors were used to formulate the search algorithms applied to the literature search.

### **Inclusion and exclusion criteria**

Though reality criteria are mainly applied in judicial contexts to ensure a victim's testimony is admitted as valid evidence, a review of the literature reveals they have been also applied to both witnesses and offenders so both populations were included as the studies were numerically sufficient for performing a meta-analysis. The concept of adult in the judicial context is associated with being 18 years of age, and the vast majority of studies endorsed this legal age; notwithstanding, in a few studies the legal age was set at 17 years. Since this difference in age has no effect on the capacity to give testimony either on cognitive or legal grounds, the studies with 17-year-old adult populations were included. The inclusion criteria for primary studies were that the effect sizes of the reality criteria analysed for discriminating between truthful and fabricated statements were reported, or in their absence, the statistical data allowing for them to be computed, including studies with errors in data analysis that nonetheless enabled the effect sizes to be computed.

The exclusion criteria were data derived from a unit of analysis which was not the statement, or when two CBCA criteria were combined into one new criterion (failing the 'mutual' exclusion requirement for creating methodic categorical systems). As for the additional criteria, data that were not formulated as additional to CBCA or were specific to only one context were excluded. Likewise, the *duplicate publication of data* was eliminated, but not the piecemeal (independent data).

Finally, 39 primary studies fulfilling the inclusion and exclusion criteria were selected. Total CBCA score was calculated using 31 effect sizes, whereas as for the individual criteria, the effect sizes ranged from 5 for criteria 10 and 19, to 35 for criteria 3 and 8.

### **Procedure**

The procedure observed the stages in meta-analysis by Botella and Gambara (2006). Having performed the literature search and selected the studies for the present meta-analysis, these were coded according to variables that have been found to have a moderating role i.e., previous studies (Fariña, Arce, & Real, 1994; Höfer et al., 1993; Raskin et al., 1991; Volbert & Steller, 2014; Vrij, 2005); previous meta-analysis with a child population (Amado et al., 2015); the research paradigm (field vs. experimental studies) under the US law of precedence (Daubert v. Merrell Dow Pharmaceuticals, 1993); compliance with the Daubert standard publication criterion (DSPC) i.e., peer-reviewed journals for evidence to be admitted as

scientifically valid legal evidence; the lie coaching condition in reality criteria; and the version of the categorical system (full reality criteria vs. 14-criteria version). Having applied a procedure of successive approximations for the coding of the primary studies (Fariña, Arce, & Novo, 2002), the following moderators were detected: status of the declarant i.e., testimony target (victim, offender, or witness); event target (self-experienced events or video-observed events/witness), judicial context i.e., case type.

As some researchers had renamed the original criteria (Steller & Köhnken, 1989), a *Thurstone style* evaluation was used consisting of 10 judges who evaluated the degree of overlapping between the original and reformulated criterion. When the interval between Q1 and Q3 was within the region of criteria independence it was considered additional criteria, whereas when it was in the region of dependence with the original, it was considered original criteria.

The coding of the studies and moderators carried out by two independent researchers showed total coincidence ( $\kappa = 1$ ).

### Data Analysis

The effect sizes were taken directly from the primary studies when these were disclosed, or the effect size  $d$  was computed using the means, and standard deviations/ standard error of the mean (Cohen's  $d$  when  $N_1 = N_2$  and Glass's  $\Delta$  when  $N_1 \neq N_2$ ), the  $t$  value, or the  $F$  value. When the results were expressed as proportions the effect size  $\delta$  (Hedges & Olkin, 1985) was equivalent to Cohen's  $d$ , whereas when they were expressed in 2X2 contingency tables, the phi obtained was transformed into Cohen's  $d$ .

The meta-analysis was performed in accordance with the procedure of Hunter and Schmidt (2015), the unit of analysis ( $n$ ) was the number of statements, the effect sizes were weighted for sample size i.e., the number of statements ( $d_w$ ), and effect sizes were corrected for criterion reliability ( $\delta$ ).

The differences between effect sizes were estimated using the difference between correlations ( $q$  statistic; Cohen, 1988), by transforming the effect sizes into correlations. In the study of moderators the average criteria for each moderator was computed.

In order to estimate the practical utility of the results of the meta-analysis in forensic settings, three recommended statistics were employed (Amado et al., 2015):  $U_1$ , the Binomial Effect Size Display (BESD), and the Probability of Superiority (PS).

### Criterion reliability

Not all of the primary studies provided data on inter-rater reliability, or agreement for the reality criteria and for the total CBCA score. Moreover, the informed reliability coefficients varied among studies, and in some studies, several were reported, in which case those approximating the results obtained by Anson, Golding and Gully (1993) and Horowitz et al. (1997) were taken. Owing to the lack of data on coding reliability in studies on specific criteria, average reliability was estimated for the criteria and for the total CBCA score, bearing in mind that reliability is different for the criteria than for the instrument (Horowitz et al., 1997). Reliability was estimated on the basis of reliability coefficients, since agreement indexes do not measure reliability. Thus, on the basis of 172 reliability coefficients of CBCA criteria in the primary studies, the average reliability for CBCA reality criteria was  $r = .61$  ( $EEM = .020$ ,  $95\%CI = 0.57, 0.65$ ); and for the total CBCA score the Spearman-Brown

prediction formula obtained an  $r = .97$ . Moreover, the average reliability for the proposed additional reality criteria was calculated using 7 reliability coefficients with an  $r$  of  $.74$  ( $EEM = 0.041$ ,  $95\%CI = 0.66, 0.82$ ). The low average reliability observed was sometimes considered as a methodological weakness of the system. Nevertheless, this potential methodological deficiency is corrected for criteria unreliability in Hunter and Schmidt's (2015) meta-analytical procedure.

## Results<sup>1</sup>

### Study of outliers

An analysis and initial control of outliers was carried out in each of the reality criteria, and the total CBCA score and conditions. The criterion chosen was the  $\pm 3 * IQR$  (extreme cases) of the simple size weighted mean effect size, given that the results of more conservative criteria such as  $\pm 1.5 * IQR$  or  $\pm 2SD$ , eliminated more than 10% of the effect sizes, indicating they were more probably moderators than outliers (Tukey, 1960).

### Global meta-analysis of reality criteria in adults

The results (Table 2) show a positive (between criteria presence and statement reality), and significant (when the confidence interval had no zero, indicating the effect size was significant) mean true effect size ( $\delta$ ) for the CBCA reality criteria, with the exception of 'self-deprecation' and 'pardoning the perpetrator' criteria, and the total CBCA score. Nevertheless, these results are not generalizable (criteria 10 and 19 were affected by a second order sampling error, so the results were invalid for this estimate) to future samples (when the credibility interval had zero, indicating the effect size was not generalizable to 80% of other samples). For the additional criteria (Höfer et al., 1993), the meta-analysis revealed a positive, significant and generalizable mean true effect size for 'reporting style' and 'display of insecurities' criteria. The mean true effect size for the 'repetitions' criterion was *negative* and significant, confirming it was not reality criteria, but not generalizable. As for the 'providing reasons for lack of memory' and 'clichés' criteria, a non-significant mean true effect size was found. The criteria repetitions and clichés were related to fabricated events, that is, they were not reality criteria in themselves, so they were not included in successive analyses. The 75% rule and the credibility interval (Hunter & Schmidt, 2015) warranted the study of moderators.

---

<sup>1</sup> Additional results and resources at [http://www.researchgate.net/profile/Ramon\\_Arce](http://www.researchgate.net/profile/Ramon_Arce)

Table 2. Results of Global Meta-analysis for Individual Reality Criteria and Total CBCA Score, and Additional Criteria

CBCA Criterion	k	n	d <sub>w</sub>	SD <sub>d</sub>	SD <sub>pre</sub>	SD <sub>res</sub>	δ	SD <sub>δ</sub>	%Var	95% CI <sub>d</sub>	80% CV <sub>δ</sub>
1. Logical structure	30	2,265	0.48	0.6990	0.2503	0.6527	0.62	0.8493	13	0.40, 0.56	-0.46, 1.71
2. Unstructured production	27	1,987	0.53	0.9241	0.2570	0.8876	0.69	1.1551	8	0.45, 0.61	-0.79, 2.17
3. Quantity of details	35	2,714	0.55	0.8294	0.2529	0.7899	0.71	1.0279	9	0.47, 0.63	-0.60, 2.03
4. Contextual embedding	29	2,137	0.19	0.6169	0.2372	0.5868	0.24	0.7411	15	0.11, 0.27	-0.70, 1.19
5. Description of interactions	29	2,243	0.27	0.3742	0.2349	0.2912	0.36	0.3790	39	0.19, 0.35	-0.13, 0.84
6. Reproduction conversations	34	2,528	0.34	0.4990	0.1780	0.4662	0.44	0.6067	13	0.26, 0.42	-0.33, 1.22
7. Unexpected complications	29	1,956	0.25	0.3788	0.2498	0.2847	0.32	0.3705	43	0.17, 0.33	-0.15, 0.79
8. Unusual details	35	2,441	0.31	0.6532	0.2489	0.6039	0.41	0.7859	14	0.23, 0.39	-0.59, 1.42
9. Superfluous details	27	1,863	0.14	0.5676	0.2437	0.5126	0.18	0.6670	18	0.04, 0.24	-0.67, 1.04
10. Details misunderstood	5	376	0.22	0.1208	0.2357	0.0000	0.28	0.0000	100	0.02, 0.42	0.28
11. External associations	22	1,612	0.26	0.4781	0.2405	0.3268	0.34	0.5376	25	0.16, 0.36	-0.35, 1.02
12. Subjective mental state	28	2,170	0.18	0.4843	0.2312	0.4256	0.23	0.5538	23	0.10, 0.26	-0.47, 0.94
13. Perpetrator's mental state	31	2,232	0.09	0.6212	0.2376	0.5741	0.11	0.7470	15	0.01, 0.17	-0.84, 1.07
14. Spontaneous corrections	29	1,842	0.16	0.5276	0.2545	0.4622	0.20	0.6014	23	0.06, 0.26	-0.56, 0.97
15. Admitting lack of memory	34	2,305	0.25	0.3823	0.2494	0.2897	0.32	0.3770	42	0.17, 0.33	-0.16, 0.80
16. Doubts one's testimony	26	1,755	0.20	0.4521	0.2478	0.3781	0.26	0.4919	30	0.10, 0.30	-0.37, 0.89
17. Self-deprecation	13	948	0.04	0.4629	0.2354	0.3985	0.05	0.5186	26	-0.08, 0.16	-0.61, 0.71
18. Pardonning the perpetrator	8	680	-0.02	0.2796	0.2178	0.1753	-0.02	0.2281	61	-0.18, 0.14	-0.31, 0.27
19. Details characteristics offence	5	562	0.28	0.1894	0.1966	0.0000	0.36	0.0000	100	0.12, 0.44	0.36
<b>TOTAL CBCA SCORE</b>	31	2,124	0.55	0.6759	0.2475	0.6290	0.56	0.6386	13	0.47, 0.63	-0.25, 1.37
<b>AVERAGE (original criteria)</b>	46	3,223	0.25	0.5032	0.2368	0.4269	0.33	0.5614	32	0.17, 0.33	-0.39, 1.05
<b>Additional Criteria</b>											
Reporting style	3	357	0.41	0.2030	0.1874	0.0781	0.48	0.0909	85	0.20, 0.63	0.36, 0.59
Display insecurities	3	297	0.67	0.5540	0.2111	0.5122	0.78	0.5965	14	0.43, 0.90	0.01, 1.54
Providing reasons lack memory	4	447	0.15	0.2877	0.1902	0.2158	0.18	0.2514	44	-0.03 0.33	-0.14, 0.50
Cliches	3	267	-0.18	0.5145	0.2134	0.4682	-0.21	0.5452	17	-0.41, 0.05	-0.90, 0.49
Repetitions	4	417	-0.47	0.5851	0.2011	0.5494	-0.54	0.6399	12	-0.67, -0.27	-1.36, 0.27
<b>AVERAGE (original+additional)</b>	46	3,223	0.27	0.4821	0.2313	0.4053	0.34	0.5275	34	0.19, 0.35	-0.33, 1.01

Note. k = number of studies; n = total sample size; d<sub>w</sub> = effect size weighted for sample size; SD<sub>d</sub> = observed standard deviation of d; SD<sub>pre</sub> = standard deviations of observed d-values corrected from all artifacts; SD<sub>res</sub> = standard deviation of observed d-values after removal of variance due to all artifacts; δ = effect size corrected for criterion reliability; SD<sub>δ</sub> = standard deviation of δ; %Var = variance accounted for by artifactual errors; 95% CI<sub>d</sub> = 95% confidence interval for d; 80% CV<sub>δ</sub> = 80% credibility interval for δ.

### Study of moderators

The study of moderators (criteria average as dependent variable; Table 3) showed a positive and significant mean true effect size, but not generalizable, in all of the moderators analysed. As for the magnitude of the effect sizes, excluding the witness condition with a medium effect size ( $\delta > 0.50$ ), all were small ( $0.20 > \delta < 0.50$ ). Arce and Fariña (2009) have suggested (and designed) the specifications of categorical systems based on bottom-up rather than 'top-down' procedures to ensure only categories that effectively discriminate between memories of experienced events and fabricated memories form part of the system. This maximizes the efficacy of the resulting categorical system by eliminating the noise produced by non-discriminating 'top-down' categories. Thus, the meta-analyses were repeated with the categories of content analysis with significant effect size i.e., the confidence interval for  $d$  did not contain zero. The results (Table 3) revealed a significant increase in the effect size of field studies,  $q_c = .119$ ,  $p < .05$  (one-tailed; a larger effect size was expected with significant criteria), thus the effect size was significantly larger with significant criteria. Moreover, for significant criteria, the results (not all of the reality criteria were generalizable) became generalizable (the credibility interval had no zero). As for the experimental studies on the remaining moderators, the results did not corroborate the Hypothesis as the reality categories had been initially or subsequently screened to eliminate the non-significant ones.

The meta-analytical technique does not take into account the theoretical foundations or the reliability of the studies included in the original theories, that is, all of the studies on categories of reality are included. Moreover, the experimental designs of studies on witnesses are not really on witnesses of self-experienced events, but on non-self-experienced events i.e., video-observed events (watched on video, not involving self-experienced events) that do not fulfil the original theory hypothesizing that reality criteria discern between memories of self-experienced real-life events and fabricated or fictitious memories. Only one of the studies on witnesses involved self-experienced events (Gödert, Gamer, Rill, & Vossel, 2005), and for the total reality criteria, were found to discriminate significantly real witness from real offenders giving false memory,  $d = 0.59$ ,  $1-\beta = .78$ , and from uninvolved participants,  $d = 0.83$ ,  $1-\beta = .96$ . Nevertheless, reality criteria also discriminated between both memories of video-observed events and fabricated events. The only study (Lee, Klaver, & Hart, 2008) comparing memories of self-experienced events (truth condition) and video-observed events (lie condition) found CBCA reality criteria, and the total CBCA score discriminated significantly between both memories in line with the Undeutsch Hypothesis.



Table 3. Results of the Meta-analysis of Moderators

MODERATOR	k	n	d <sub>w</sub>	SD <sub>d</sub>	SD <sub>pre</sub>	SD <sub>res</sub>	δ	SD <sub>δ</sub>	%Var	95% CI <sub>d</sub>	80% CV <sub>δ</sub>
CBCA significant criteria (17)	46	3,223	0.27	0.5187	0.2380	0.4433	0.36	0.5835	31	0.19, 0.35-0.39, 1.11	0.22, 0.34-0.47, 1.19
14-criteria version	45	3,143	0.28	0.5567	0.2394	0.4906	0.36	0.6465	25		
DAUBERT STANDARD PUBLICATION CRITERION											
All criteria (22)	35	2,256	0.20	0.4575	0.2407	0.3733	0.26	0.4786	39	0.12, 0.28	-0.35, 0.87
SELF-EXPERIENCED EVENTS											
All criteria (22)	34	2,277	0.26	0.4647	0.2371	0.3879	0.33	0.5022	40	0.18, 0.34	-0.31, 0.97
NON SELF-EXPERIENCED EVENTS (WITNESS)											
All criteria (13)	11	625	0.39	0.5835	0.2707	0.5032	0.51	0.6548	65	0.23, 0.55	-0.33, 1.35
OFFENDERS											
All criteria (21)	11	1,067	0.27	0.4662	0.2024	0.3743	0.35	0.4975	41	0.15, 0.39	-0.29, 0.99
VICTIMS											
All criteria (18)	11	840	0.27	0.4781	0.2355	0.4012	0.35	0.5221	35	0.13, 0.41	-0.32, 1.02
FIELD STUDIES											
All field studies (18)	6	422	0.34	0.4948	0.2385	0.4153	0.45	0.5404	35	0.14, 0.54-0.24, 1.14	
Significant criteria (10) <sup>a</sup>	6	422	0.53	0.4774	0.2458	0.3834	0.69	0.4989	42	0.33, 0.73	0.05, 1.33
SEXUAL AND IPV FIELD STUDIES											
All criteria (17) <sup>b</sup>	5	263	0.67	0.3587	0.2871	0.1957	0.87	0.2459	72	0.41, 0.92	0.55, 1.18
Significant criteria (15) <sup>c</sup>	5	263	0.74	0.3654	0.2892	0.2134	0.96	0.2478	72	0.48, 0.99	0.64, 1.28
EXPERIMENTAL STUDIES											
All criteria (22)	39	2,721	0.25	0.4497	0.2336	0.3934	0.32	0.4933	37	0.17, 0.33	-0.31, 0.95

Note. <sup>a</sup>Significant criteria (CBCA criteria, as for additional criteria, studies were insufficient); 1-3, 5-8, 11, 12 and 19; <sup>b</sup>Significant criteria (CBCA criteria): 1-9, 11-18; <sup>c</sup>Significant criteria (CBCA criteria): 1-9, 11-12, 14-17.

The high observed variability in effect sizes in field studies, which was mostly due to one study alone, suggested differences in experimental design (the crime context in this study was found to be different to the other studies). As the effect of context has been hypothesized (Köhnken, 1996; Volbert & Steller, 2014), and found (Arce, Fariña, & Vilariño, 2010; Vilariño et al. 2011) to mediate the discriminating efficacy of reality categories, the meta-analysis was repeated in field studies on sexual offences and intimate partner violence (IPV) cases (crimes committed in the privacy of one's home according to the categorization of Arce and Fariña, 2005). The results showed a positive, significant and generalizable (not generalizable in all field studies) mean true effect size for studies under this condition. Moreover, the magnitude of the effect sizes were significantly larger in sexual offences and IPV cases than in all field studies in all the reality criteria (0.45 for all field studies vs. 0.87 for sexual offences and IPV cases),  $q_c = .199$ ,  $p < .01$  (one-tailed; a higher effect size was expected in specific criminal contexts), and in the significant criteria,  $q_c = .168$ ,  $p < .05$  (0.69 vs. 0.96). Likewise, reality criteria were significantly more efficacious,  $q_c = .2622$ ,  $p < .01$ , in sexual offences and IPV cases than in all other types of cases (0.32 vs. 0.87).

Results (meta-analysis could not be performed because *ks* and *ns* were insufficient and research designs incomparable) for the comparison between statements of participants instructed to lie (lie coaching condition) with truthful statements were inconclusive<sup>2</sup> in relation to the effectiveness of reality criteria to discriminate between truthful and false statements.

## Discussion

The following conclusions may be drawn from the results of this study. First, the results confirmed the Undeutsch Hypothesis, that is, reality criteria discriminated between memories of self-experienced and fabricated events [File Drawer Analysis (FDA): to bring down this hypothesis to a trivial effect (McNatt, 2000), .05, for the average of the CBCA criteria, it would be necessary 184 studies with null effect; Hunter & Schmidt, 2015. It is unlikely to happen]. Besides fulfilling the DSPC, this Hypothesis was also valid for memories of victims/claimants and offenders (for witness of self-experienced events further research is required); and robust in both experimental studies (high internal validity), and field studies (high external validity). Notwithstanding, the reality criteria also discriminated between memories of video-observed events i.e., non-self-experienced events, and fabricated events for which the Hypothesis was not formulated, and research findings are inconclusive as to the validity of the Hypothesis with lie coached subjects. Second, though the results validated CBCA as a categorical system based on the Undeutsch Hypothesis, neither were all of the criteria validated, nor were they generalizable, and some even contradicted the Hypothesis. Thus, these criteria can be used neither in all types of contexts, nor indiscriminately. Both versions of the CBCA (all criteria or 14 criteria) were exactly the same ( $\delta = 0.36$ ) in discriminating between memories of self-experienced and fabricated events. Though the results open the door to the inclusion of new reality criteria, additional criteria have been proposed that fail to fulfil the Undeutsch Hypothesis (significant negative effect sizes i.e., not

<sup>2</sup> Conclusions in the primary studies about non-significant results are inconclusive as the statistical power,  $1 - \beta < .80$ , is insufficient to conclude ( $d = -0.44$ ,  $1 - \beta = .41$ , Bogaard, Meijer, & Vrij, 2013;  $d = 0.37$ ,  $1 - \beta = .26$ , Vrij, Akehurst, Soukara, & Bull, 2002;  $d = 0.11$ ,  $1 - \beta = .06$ , Vrij, Kneller, & Mann, 2000).

reality criteria), so they cannot be included in the CBCA. Third, in field studies the discriminating power of reality criteria was significantly higher in sexual offences and IPV cases (FDA: to bring the results in sexual offences and IPV cases down to a trivial effect, it would be necessary 62 and 69 studies with null effect for all criteria and significant criteria, respectively. It is unlikely to occur) in comparison to other types of contexts (FDA: to reduce the efficacy of the reality criteria to discriminate between real and fabricated memories in any context of field studies to a trivial effect it would be necessary 35 studies with null effect. It is unlikely to happen). Succinctly, the areas of both populations do not overlap in 54% ( $U1 = 0.54$ ), that is, they were totally independent, thus the efficacy of the reality criteria in discriminating between memories of self-experienced and fabricated events in sexual and IPV cases was total in 54% of the evaluations of credibility. Moreover, 75% of statements of self-experienced events contained more reality criteria than fabricated events (probability of superiority,  $PS = 0.75$ ), the probability of false positives was 28% (BESD). These results were highly robust i.e., not only establishing a positive and significant relation between reality criteria and true statements, but were also generalizable to all types of sexual offences and IPV cases, and were homogeneous (i.e., subject to little variability since the correlation between the effect sizes was .72).

As for the implications for forensic practice, the results of the present meta-analysis reveal that the reality criteria were statistically effective for discriminating between memories of self-experienced and fabricated events, but this does not imply they are directly generalizable to forensic practice. Even under the best discriminating conditions i.e., field studies in sexual and IPV cases, the probability of false positives may reach .22, whilst this probability must be zero in forensic settings (Arce, Fariña, & Fraga, 2000). In general, only significant reality criteria i.e., scientifically attested evidence, were admissible for forensic practice (see note of Table 3), since the results were generalizable, whereas for all criteria they were not. However, as the credibility interval lower limit was 0.05, the practical utility of these categories was almost negligible ( $PS=.51$ ), that is, in only 51% of true statements there were more reality criteria than in false statements, and under what specific conditions this contingency occurred remains unknown. However, the credibility interval lower limit of the reality criteria applied to cases of sexual offences and IPV, which were also generalizable both in terms of all the criteria and the significant criteria, was larger,  $PS = .73$  and  $.75$  (Hedges and Olkin's  $\delta = 0.59$  and  $0.65$ , test value =  $.51$ ), for all the reality criteria and the significant criteria, respectively. However, these conclusions are not directly applicable to forensic practice as the decision criteria which in the forensic context must be the 'strict decision criterion' in which a type II error (classify a false statement as true) is not admissible i.e., must be equal to zero. Regarding the strict decision criterion, Arce et al. (2010) found up to 13 CBCA reality criteria in fabricated statements of IPV cases, which means that at least 14 reality criteria would have to be detected in a statement to conclude that the testimony was true, with a correct classification of true positives (true statements classified as such) of 36%. Succinctly, the CBCA reality criteria were a poor tool for assigning the credibility of IPV victim testimony. Thus, to enhance efficacy, CBCA reality criteria must be complemented with additional criteria. In this line, Arce and Fariña (2009), Vilariño, (2010) and Vilariño et al. (2011) combined CBCA and SRA criteria, memory attributes, and additional reality criteria specific to IPV cases derived from real statements (judicial judgements as ground truth), to create and validate a categorical system specific for IPV cases, including sexual offences, with a strict decision criterion to reduce the rate of false negatives to 2%. In any way, only results with a strict decision criterion can be translated into forensic practice.

In terms of future research, the results of the present meta-analysis underscored the need for further studies with experimental designs assessing the efficacy of reality criteria in discriminating between memories of self-experienced events and video witnessed non-self-experienced events; between self-experienced witnessed events vs. fabricated events; between memories of participants coached to lie and honest; and research driven to find new reality categories (bottom-up), mainly for a specific context i.e., crime victimization.

This meta-analysis is subject to the following limitations. First, previous publications have biased the results in that the non-significant results or predictably inefficacious categories were eliminated (favouring the validation of the Undeutsch Hypothesis). Second, the feigning methodology (experimental studies) had no proven external validity (Sarwar, Allwood, & Innes-Ker, 2014), but only 'face validity' (Konecni & Ebbesen, 1992). Third, for some experimental literature, statements are insufficient material for reality content analysis (Köhnken, 2004), which favours the rejection of the Undeutsch Hypothesis. Fourth, there was no control on the effects of the interviewer on the contents of the statement, or on the reliability of the interviews, which were often carried out by poorly trained interviewers. Fifth, few studies comply with SVA standards that are a requirement for applying CBCA. Sixth, the results of some meta-analysis may be subject to a degree of variability, given that  $N_s < 400$ , did not guarantee stability in sample estimates (Hunter & Schmidt, 2015). Seventh, primary studies did not estimate the reliability of the codings, thus results' reliability is uncertainty.

### References

- \*Akehurst, L., Easton, S., Fullar, E., Drane, G., Kuzmin, K., & Litchfield, S. (2015). An evaluation of a new tool to aid judgements of credibility in the medico-legal setting. *Legal and Criminological Psychology*. Advance online publication. doi: 10.1111/lcrp.12079
- Amado, B.G., Arce, R., & Fariña, F. (2015). Undeutsch hypothesis and Criteria Based Content Analysis: A meta-analytic review. *European Journal of Psychology Applied to Legal Context*, 7, 3-12. doi:10.1016/j.ejpal.2014.11.002
- Anson, D.A., Golding, S.L., & Gully, K.J. (1993). Child sexual abuse allegations: Reliability of Criteria-Based Content Analysis. *Law and Human Behavior*, 17, 331-341. doi:10.1007/BF01044512
- Arce, R., & Fariña, F. (2005). Peritación psicológica de la credibilidad del testimonio, la huella psíquica y la simulación: El Sistema de Evaluación Global (SEG) [Psychological evidence in court on statement credibility, psychological injury and malingering: The Global Evaluation System (GES)]. *Papeles del Psicólogo*, 26, 59-77.
- Arce, R., & Fariña, F. (2009). Evaluación psicológica forense de la credibilidad y daño psíquico en casos de violencia de género mediante el Sistema de Evaluación Global. In F. Fariña, R. Arce, & G. Buela-Casal (Eds.), *Violencia de género. Tratado psicológico y legal* (pp. 147-168). Madrid: Biblioteca Nueva.
- Arce, R., & Fariña, F. (2012). Psicología social aplicada al ámbito jurídico. In A.V. Arias, J.F. Morales, E. Nouvilas, & J.L. Martínez (Eds.), *Psicología social aplicada* (pp. 157-182). Madrid, Spain: Panamericana.

- Arce, R., Fariña, F., & Fraga, A. (2000). Género y formación de juicios en un caso de violación [Gender and juror judgment making in a case of rape]. *Psicothema, 12*, 623-628.
- \*Arce, R., Fariña, F., & Vilariño, M. (2010). Contraste de la efectividad del CBCA en la evaluación de la credibilidad en casos de violencia de género [Contrasting the efficiency of the CBCA in the assessment of credibility in violence against women cases]. *Intervención Psicosocial, 19*, 109-119. doi:10.5093/in2010v19n2a2
- \*Beaulieu-Prévost, D. (2001). *Analyse de validité de la déclaration (SVA), mensonge et faux souvenirs: Validité et efficacité chez les adultes*. (Doctoral dissertation). Retrieved from ProQuest Dissertations & Theses Global. (Order No. MQ60609)
- \*Bensi, L., Gambetti, E., Nori, R., & Giusberti, F. (2009). Discerning truth from deception: The sincere witness profile. *European Journal of Psychology Applied to Legal Context, 1*, 101-121.
- Berliner, L., & Conte, J.R. (1993). Sexual abuse evaluation: Conceptual and empirical obstacles. *Child Abuse and Neglect, 17*, 111-125. doi:10.1016/0145-2134(93)90012-T
- \*Biland, C., Py, J., & Rimboud, S. (1999). Evaluer la sincérité d'un témoin grâce à trois techniques d'analyse, verbales et non verbales [Three verbal or nonverbal techniques for evaluating sincerity]. *European Review of Applied Psychology, 49*, 115-122.
- \*Blandón-Gitlin, I., Pezdek, K., Lindsay, D.S., & Hagen, L. (2009). Criteria-Based Content Analysis of true and suggested accounts of events. *Applied Cognitive Psychology, 23*, 901-917. doi:10.1002/acp.1504
- \*Bogaard, G., Meijer, E.H., & Vrij, A. (2013). Using an example statement increases information but does not increase accuracy of CBCA, RM, and SCAN. *Journal of Investigative Psychology and Offender Profiling, 11*, 151-163. doi:10.1002/jip.1409
- Botella, J. & Gambará, H. (2006). Doing and reporting a meta-analysis. *International Journal of Clinical & Health Psychology, 6*, 425-440.
- \*Caso, L., Vrij, A., Mann, S., & de Leo, G. (2006). Deceptive responses: The impact of verbal and non-verbal countermeasures. *Legal and Criminological Psychology, 11*, 99-111. doi:10.1348/135532505X49936
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: LEA.
- \*Critchlow, N. (2011). *Applying Criteria Based Content Analysis to assessing the veracity of rape statements* (Unpublished doctoral dissertation). Manchester Metropolitan University, Manchester, UK.
- \*Critchlow, N. (2011). [A field validation of CBCA when assessing authentic police rape statements: evidence for discriminant validity to prescribe veracity to adult narrative]. Unpublished raw data.
- \*Dana-Kirby, L. (1997). *Discerning truth from deception: Is Criteria-Based Content Analysis effective with adult statements?* (Unpublished doctoral thesis). University of Oregon, Oregon.
- Daubert v. Merrell Dow Pharmaceuticals, 509 U.S. 579 (1993).
- \*Evans, J., Michael, S.W., Meissner, C.A., & Brandon, S.E. (2013). Validating a new assessment method for deception detection: Introducing a psychologically based credibility assessment tool. *Journal of Applied Research in Memory and Cognition, 2*, 33-41. doi:10.1016/j.jarmac.2013.02.002
- Fariña, F., Arce, R., & Novo, M. (2002). Heurístico de anclaje en las decisiones judiciales [Anchorage in judicial decision making]. *Psicothema, 14*, 39-46.

- Fariña, F., Arce, R., & Real, S. (1994). Ruedas de identificación: De la simulación y la realidad [Linepus: A comparison of high-fidelity research and research in a real context]. *Psicothema*, *6*, 395-402.
- \*Gödert, H.W., Gamer, M., Rill, H.G., & Vossel, G. (2005). Statement Validity Assessment: Inter-rater reliability of Criteria-Based Content Analysis in the mock-crime paradigm. *Legal and Criminological Psychology*, *10*, 225-245. doi:10.1348/135532505X52680
- \*Godoy, V., & Higuera, L. (2008). El análisis de contenido basado en criterios (CBCA) y la entrevista cognitiva aplicados a la credibilidad del testimonio en adultos. In F. Rodríguez, C. Bringas, F. Fariña, R. Arce, & A. Bernardo (Eds.), *Psicología Jurídica: Entorno judicial y delincuencia* (pp. 117-125). Retrieved from <http://gip.uniovi.es/T5EJD.pdf>
- Hans, V.P., & Vidmar, N. (1986). *Judging the jury*. New York: Plenum Press.
- Hedges, L.V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.
- Höfer, E., Köhnken, G., Hanewinkel, R., & Bruhn, C. (1993). *Diagnostik und attribution von glaubwürdigkeit*. Unpublished final report. University of Kiel, Germany.
- \*Honts, C.R., & Devitt, M.K. (1993). *Credibility Assessment of Verbatim Statements (CAVS)*. Retrieved from <http://www.dtic.mil/dtic/tr/fulltext/u2/a271575.pdf>
- Horowitz, S.W., Lamb, M.E., Esplin, P.W., Boychuk, T.D., Krispin, O., & Reiter-Lavery, L. (1997). Reliability of criteria-based content analysis of child witness statements. *Legal and Criminological Psychology*, *2*, 11-21. doi:10.1111/j.2044-8333.1997.tb00329.x
- Hunter, J.E., & Schmidt, F.L. (2015). *Methods of meta-analysis: Correcting error and bias in research findings*. Newbury Park, CA: Sage.
- \*Johnston, S., Candelier, A., Powers-Green, D., & Rahmani, S. (2014). Attributes of truthful versus deceitful statements in the evaluation of accused child molesters. *Sage Open*, *4*(3), 1-10. doi: 10.1177/2158244014548849
- \*Juárez, J.R., Mateu, A., & Sala, E. (2007). *Criterios de evaluación de la credibilidad en las denuncias de violencia de género*. Retrieved from <http://justicia.gencat.cat/web/.content/documents/arxiu/sc-3-143-07-cas.pdf>
- Köhnken, G. (1996). Social psychology and the law. In G.R. Semin & K. Fiedler (Eds.), *Applied social psychology* (pp. 257-282). Thousand Oaks, CA: Sage.
- Köhnken, G. (2004). Statement Validity Analysis and the detection of the truth. In P. A. Granhag & L. A. Strömwall (Eds.), *The detection of deception in forensic contexts* (pp. 41-63). Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9780511490071.003
- \*Köhnken, G., Schimossek, E., Aschermann, E., & Höfer, E. (1995). The cognitive interview and the assessment of the credibility of adults' statements. *Journal of Applied Psychology*, *80*, 671-684. doi:10.1037/0021-9010.80.6.671
- Konecni, V.J., & Ebbesen, E.B. (1992). Methodological issues on legal decision-making, with special reference to experimental simulations. In F. Lösel, D. Bender, & T. Bliesener (Eds.), *Psychology and law. International perspectives* (pp. 413-423). Berlin, Germany: Walter de Gruyter
- \*Leal, S., Vrij, A., Warmelink, L., Vernham, Z., & Fisher, R.P. (2015). You cannot hide your telephone lies: Providing a model statement as an aid to detect deception in insurance telephone calls. *Legal and Criminological Psychology*, *20*, 129-146. doi:10.1111/lcrp.12017
- \*Lee, Z., Klaver, J.R., & Hart, S.D. (2008). Psychopathy and verbal indicators of deception in offenders. *Psychology, Crime & Law*, *14*, 73-84. doi:10.1080/10683160701423738

- McNatt, D.B. (2000). Ancient Pygmalion joins contemporary management: A meta-analysis of the result. *Journal of Applied Psychology*, 85, 314-322. doi:10.1037/0021-9010.85.2.314
- \*Merckelbach, H. (2004). Telling a good story: Fantasy proneness and the quality of fabricated memories. *Personality and Individual Differences*, 37, 1371-1382. doi:10.1016/j.paid.2004.01.007
- Novo, M., & Seijo, D. (2010). Judicial judgement-making and legal criteria of testimonial credibility. *The European Journal of Psychology Applied to Legal Context*, 2, 91-115.
- \*Porter, S., & Yuille, J.C. (1996). The language of deceit: An investigation of the verbal clues to deception in the interrogation context. *Law and Human Behavior*, 20, 443-458. doi:10.1007/BF01498980
- \*Porter, S., Yuille, J.C., & Lehman, D.R. (1999). The nature of real, implanted, and fabricated memories for emotional childhood events: Implications for the recovered memory debate. *Law and Human Behavior*, 23, 517-537. doi:10.1023/A:1022344128649
- Raskin, D.C., Esplin, F.W., & Horowitz, S. (1991). *Investigative interviews and assessment of children in sexual abuse cases*. Unpublished manuscript, Department of Psychology, University of Utah, Utah.
- \*Rassin, E., & van-der-Sleen, J. (2005). Characteristics of true versus false allegations of sexual offences. *Psychological Reports*, 97, 589-598. doi:10.2466/pr0.97.2.589-598
- Sarwar, F., Allwood, C. M., & Innes-Ker, A. (2014). Effects of different types of forensic information on eyewitness' memory and confidence accuracy. *European Journal of Psychology Applied to Legal Context*, 6, 17-27. doi: 10.5093/ejpalc2014a3
- \*Schelleman-Offermans, K., & Merckelbach, H. (2010). Fantasy proneness as a confounder of verbal lie detection tools. *Journal of Investigative Psychology and Offender Profiling*, 7, 247-260. doi:10.1002/jip.121
- \*Sporer, S.L. (1997). The less travelled road to truth: Verbal cues in deception detection in accounts of fabricated and self-experienced events. *Applied Cognitive Psychology*, 11, 373-397. doi:10.1002/(SICI)1099-0720(199710)11:5<373::AID-ACP461>3.0.CO;2-0
- Steller, M., & Böhm, C. (2006). Cincuenta años de jurisprudencia del Tribunal Federal Supremo alemán sobre la psicología del testimonio. Balance y perspectiva. In T. Fabian, C. Böhm, & J. Romero (Eds.), *Nuevos caminos y conceptos en la psicología jurídica* (pp. 53-67). Münster, Germany: LIT Verlag.
- Steller, M., & Köhnken, G. (1989). Criteria-Based Content Analysis. In D.C. Raskin (Ed.), *Psychological methods in criminal investigation and evidence* (pp. 217-245). New York: Springer-Verlag.
- \*Ternes, M. (2009). *Verbal credibility assessment of incarcerated violent offenders' memory reports* (Unpublished doctoral thesis). University of British Columbia, Vancouver.
- Tukey, J.W. (1960). A survey of sampling from contaminated distributions. In I. Olkin, J.G. Ghurye, W. Hoeffding, W.G. Madow, & H. Mann (Eds.), *Contributions to probability and statistics* (pp. 448-485). Stanford, CA: Stanford University Press.
- Undeutsch, U. (1967). Beurteilung der glaubhaftigkeit von aussagen. In U. Undeutsch (Ed.), *Handbuch der psychologie, Vol. 11: Forensische psychologie* (pp. 26-181). Göttingen, Germany: Hogrefe.
- \*Vilariño, M. (2010). *¿Es posible discriminar declaraciones reales de imaginadas y huella psíquica real de simulada en casos de violencia de género?* (Doctoral thesis, Universidad de Santiago de Compostela, Spain). Retrieved from <http://hdl.handle.net/10347/2831>
- Vilariño, M., Novo, M., & Seijo, D. (2011). Estudio de la eficacia de las categorías de realidad del testimonio del Sistema de Evaluación Global (SEG) en casos de violencia de género

- [Study of the efficacy of the testimony reality categories of the Global Evaluation System (GES) in violence against women cases]. *Revista Iberoamericana de Psicología y Salud*, 2, 1-26.
- Volbert, R., & Steller, M. (2014). Is this testimony truthful, fabricated, or based on false memory? Credibility assessment 25 years after Steller and Köhnken (1989). *European Psychologist*, 19, 207-220. doi:10.1027/1016-9040/a000200
- Vrij, A. (2005). Criteria-Based Content Analysis: A qualitative review of the first 37 studies. *Psychology, Public Policy, and Law*, 11, 3-41. doi:10.1037/1076-8971.11.1.3
- Vrij, A. (2008). *Detecting lies and deceit: Pitfalls and opportunities* (2nd ed.). Chichester, UK: John Wiley and Sons
- Vrij, A., Akehurst, L., Soukara, S., & Bull, R. (2002). Will the truth come out? The effect of deception, age, status, coaching and social skills on CBCA scores. *Law and Human Behavior*, 26, 261-283. doi:10.1023/A:1015313120905
- \*Vrij, A., Akehurst, L., Soukara, R., & Bull, R. (2004). Detecting deceit via analyses of verbal and nonverbal behavior in children and adults. *Human Communication Research*, 30, 8-41. doi:10.1111/j.1468-2958.2004.tb00723.x
- \*Vrij, A., Edward, K., & Bull, R. (2001). People's insight into their own behaviour and speech content while lying. *British Journal of Psychology*, 92, 373-389. doi:10.1348/000712601162248
- \*Vrij, A., Edward, K., Roberts, K.P., & Bull, R. (2000). Detecting deceit via analysis of verbal and nonverbal behavior. *Journal of Nonverbal Behavior*, 24, 239-263. doi:10.1023/A:1006610329284
- \*Vrij, A., Evans, H., Akehurst, L., & Mann, S. (2004). Rapid judgements in assessing verbal and nonverbal cues: Their potential for deception researchers and lie detection. *Applied Cognitive Psychology*, 18, 283-296. doi:10.1002/acp.964
- \*Vrij, A., & Heaven, S. (1999). Vocal and verbal indicators of deception as a function of lie complexity. *Psychology, Crime and Law*, 5, 203-215. doi:10.1080/10683169908401767
- \*Vrij, A., Kneller, W., & Mann, S. (2000). The effect of informing liars about Criteria-Based Content Analysis on their ability to deceive CBCA-raters. *Legal and Criminological Psychology*, 5, 57-70. doi:10.1348/135532500167976
- \*Vrij, A., & Mann, S., (2006). Criteria-Based Content Analysis: An empirical test of its underlying processes. *Psychology, Crime and Law*, 12, 337-349. doi:10.1080/10683160500129007
- \*Vrij, A., Mann, S., & Edward, K. (2000). I think it was a green scarf but I am not sure. Raising doubts about one's own testimony during lying and truth telling. In A. Czerederecka, T. Jaskiewicz-Obydzinska, & J. Wójcikiewicz (Eds.), *Forensic, psychology and law. Traditional questions and new ideas* (pp. 205-207). Institute of forensic research, Poland: Krakow.
- \*Vrij, A., Mann, S., Kristen, S., & Fisher, R.P. (2007). Cues to deception and ability to detect lies as a function of police interview styles. *Law and Human Behavior*, 31, 449-518. doi: 10.1007/s10979-006-9066-4
- \*Willén, R.M., & Strömwall, L.A. (2011). Offender's uncoerced false confessions: A new application of statement analysis? *Legal and Criminological Psychology*, 17, 346-359. doi:10.1111/j.2044-8333.2011.02018.x.
- \*Wojciechowski, B.W. (2014). Content analysis algorithms: An innovative and accurate approach to statement veracity assessment. *European Polygraph*, 8, 119-128. doi:0.2478/ep-2014-0010



### **8.3. PSYCHOLOGICAL INJURY IN VICTIMS OF CHILD SEXUAL ABUSE: A META-ANALYTIC REVIEW**

Bárbara G. Amado, Ramón Arce, & Andrés Herraiz

Departament of Psicología Organizacional, Jurídico-Forense y Metodología. University of Santiago de Compostela

#### **Abstract**

In order to assess the effects of child/adolescent sexual abuse (CSA/ASA) on the victim's probability of developing symptoms of depression and anxiety, to quantify injury in populational terms, to establish the probability of injury, and to determine the different effects of moderators on the severity of injury, a meta-analysis was performed. Given the abundant literature, only studies indexed in the scientific database of reference, the Web of Science, were selected. A total of 78 studies met the inclusion criteria: measured CSA/ASA victims; measured injury in terms of depression or anxiety symptoms; measured the effect size or included data for computing them; and provided a description of the sample. The results showed that CSA/ASA victims suffered significant injury, generally of a medium effect size, and generalizable; victims had 70% more probabilities of suffering from injury; and clinical diagnosis was a significantly more adequate measure of injury than symptoms. The probability of chronic injury (dysthymia) was greater than developing more severe injury i.e., major depressive disorder (MDD). In the category of anxiety disorders, injury was expressed with a higher probability in specific phobia. In terms of the victim's gender, females had significantly higher rates of developing a depressive disorder (DD) and/or an anxiety disorder (AD), quantified in a 42% and 24% over the baseline, for a DD and AD, respectively. As for the type of abuse, the meta-analysis revealed that abuse involving penetration was linked to severe injury whereas abuse with no contact was associated to less serious injury. The clinical, social, and legal implications of the results are discussed.

**Keywords:** child sexual abuse, adolescent sexual abuse, psychological injury, victimization, meta-analysis.

#### **Daño psicológico en víctimas de abuso sexual infantil: Una revision meta-analítica**

#### **Resumen**

Con el objetivo de conocer los potenciales efectos de la victimización de abuso sexual infantil/adolescente (ASI/ASA) en el desarrollo de sintomatología depresiva y ansiosa así como cuantificar, en su caso, el potencial daño en términos poblacionales, la probabilidad de manifestación de daño y el efecto diferencial de moderadores en la severidad del daño manifestado, se planificó una revisión meta-analítica. Dada la gran proliferación de literatura se seleccionaron aquellos estudios indexados en la base de datos de referencia de calidad científica, la Web of Science. 78 estudios cumplieron los criterios de inclusión: medida de la victimización de ASI/ASA; medida del daño en sintomatología depresiva o ansiosa; medida del tamaño del efecto o inclusión de datos que permitieran computarlo; y descripción de la muestra. Los resultados mostraron que la victimización de ASI/ASA conlleva a un daño

significativo de un tamaño, en general, moderado, y generalizable; que las víctimas tienen un 70% más de probabilidades de presentar daño; y que el diagnóstico clínico es una medida significativamente más adecuada del daño que la sintomatología. La probabilidad de cronificación del daño (distimia) es mayor que la de un daño más grave (depresión mayor). En la categoría de los trastornos de ansiedad, el daño se manifiesta con mayor probabilidad en fobia específica. Por el género de la víctima, las mujeres presentan una tasa significativamente mayor de desarrollo de un cuadro depresivo, cuantificado en un 42% sobre la línea base, y ansioso, cuantificado en un 24%. Por el tipo de abuso, los meta-análisis evidencian que el abuso con penetración conlleva más daño y el sin contacto un daño menor. Se discuten las implicaciones clínicas, sociales y legales de los resultados.

**Palabras clave:** abuso sexual infantil; abuso sexual adolescente; daño psicológico; victimización; meta-análisis.

### **Dano psicológico en vítimas de abuso sexual infantil: Unha revisión meta-analítica**

#### **Resumo**

Co obxecto de coñecer os potenciais efectos da vitimización de abuso sexual infantil/adolescente (ASI/ASA) no desenrolo de sintomatoloxía depresiva e ansiosa así como cuantificar, no seu caso, o potencial dano en termos poboacionais, a probabilidade de manifestación de dano e o efecto diferencial de moderadores na severidade do dano manifestado, planificouse unha revisión meta-analítica. Dado a gran proliferación de literatura seleccionáronse aqueles estudos indexados na base de datos de referencia de calidade científica, a Web of Science. Setenta e oito estudos cumpriron os criterios de inclusión: medida da vitimización de ASI/ASA; medida do dano en sintomatoloxía depresiva ou ansiosa; medida do tamaño do efecto ou inclusión de datos que permitiran computalo; e a descrición da mostra. Os resultados amosaron que a vitimización de ASI/ASA conleva un dano significativo dun tamaño do efecto, en xeral, moderado, e xeralizable; que as vítimas teñen un 70% máis de probabilidades de presentar dano; e que o diagnóstico clínico é unha medida significativamente máis adecuada do dano que a sintomatoloxía. A probabilidade de cronificación do dano (distimia) é maior que a dun dano máis grave (depresión maior). Na categoría de trastornos de ansiedades, o dano maniféstase con maior probabilidade en fobia específica. Atendendo ó xénero da vítima, as mulleres presentan unha taxa significativamente maior de desenrolo dun cadro depresivo, cuantificado nun 42% sobre a liña base, e ansioso, cuantificado nun 24%. Polo tipo de abuso, os meta-análises evidencian que o abuso con penetración conleva máis dano e o sen contacto un dano menor. Discútese as implicacións clínicas, sociais e legais dos resultados.

**Palabras chave:** abuso sexual infantil; abuso sexual adolescente; dano psicológico; vitimización; meta-análise.

## Introduction

The World Health Organization (WHO, 1999) defines child sexual abuse (CSA) as involvement of a child in sexual activity that he or she does not fully comprehend, is unable to give informed consent to, or for which the child is not developmentally prepared and cannot give consent, or that violates the laws or social taboos of society. This definition encompasses the universal criterion, the involvement of a child in sexual behaviours/activities that the child is neither physically nor mentally prepared, and who lacks the capacity to consent, as well as the legal standards specific to each country (Stoltenborgh, van Ijzendoorn, Euser, & Bakermans-Kranenburg, 2011). These authors often refer to the characteristics of the aggressor (e.g., age, family relationship), and define the typologies of abuse (e.g., contact, exhibitionism). In the scientific literature, the term child sexual abuse includes adolescent sexual abuse (ASA) (Rich, Gidycz, Warkentin, Loh, & Weiland, 2005; Schoedl et al., 2010).

A recent meta-analysis established the global prevalence of CSA/ASA at 11.8% (Stoltenborgh et al., 2011). Additionally, both epidemiological studies (WHO, 2014), and several meta-analysis (Pereda, Guilera, Forns, & Gómez-Benito, 2009; Stoltenborgh et al., 2011) have coincided in that females had CSA/ASA rates around 18% to 20% higher than the 8% prevalence rate for males.

Of the different forms of child abuse, CSA/ASA is linked to severe injury (Intebi, 1998). In fact, numerous empirical studies have established a relationship between CSA/ASA and psychological injury (e.g., major depressive disorder, dysthymia, generalized anxiety disorder, phobic disorders), which may become manifest in the short-term and/or become chronic (Jumper, 1995; Paolucci, Genuis, & Violato, 2001). Psychological injury is classified according to two broad diagnostic categories: mood disorders and depressive disorders (as termed by the DSM-IV and DSM-V, respectively), and anxiety disorders (Jumper, 1995; Maniglio, 2009; Paolucci et al., 2001). The manifestation of sequelae in symptoms or in a particular disorder was related to the variables specific to each individual.

The high CSA/ASA prevalence rates, the association between victims and mental injury, the severity and chronicity of injury, the variability in prevalence, severity, and chronicity of injury, and the contradictory results underscored the need for undertaking a meta-analysis to review the empirical data on psychological injury in CSA/ASA victims in terms of symptoms, and anxiety and depressive disorders.

Though injury associated to CSA/ASA victims is assumed to be general, certain moderators are expected to explain differences in the degree of injury. Succinctly, the prevalence and severity of injury has been reported to be different for females than for males (Jonas et al., 2011; Koenen & Widom, 2009; Stoltenborgh et al., 2011; Tolin & Foa, 2006; WHO, 2000, 2014); the type of sexual abuse suffered (i.e., contact, no contact, intercourse) (Bulik, Prescott, & Kendler, 2001; Cutajar et al., 2010; Fergusson, McLeod, & Hordwood, 2013; Jonas et al., 2011); the type of measure of psychological injury, symptoms (psychometric measure), and diagnosis of the disorders (clinical diagnosis) (Maniglio, 2010; Peleikis, Mykletum, & Dahl, 2005; Vilariño, Arce, & Fariña, 2013); and cultural aspects related to the manifestation of symptoms, and anxiety and depressive disorders (American Psychiatric Association, 2013).

Owing to the clinical, social, and legal implications of the results, the meta-analytical technique to be employed should not be constrained to standard effect sizes with their significance, generalization, and assessment of moderators, but should also quantify injury in terms of populations (i.e., assessing injury rates above the baseline) and evaluate the probability of manifesting injury, and the differential effects of moderators on the severity of injury.

## Method

### Database search of studies

The process for selecting scientific studies began with a search in the meta-search engines (i.e., Google, Yahoo, Google Scholar), which, according to the descriptors, listed between approximately 36,000 to 770,000 results. The vast majority of the search results yielded, without compromising the requirements for performing a robust meta-analysis (i.e., sufficient  $k$  and  $N$ ), a selection of studies *par excellence*. Thus, the next step was to search for studies in one of the world's leading scientific databases of reference, the Web of Science. All of the databases of scientific papers (Core Collection, Current Contents, Medline, Scielo, KCI-Korean), were searched to include not only scientific literature *par excellence*, but also cultural concepts (i.e., Latinos, Asiatic), which are referred to in the manuals for the classification of mental disorders (DSM and ICD), and may have differential effects on the symptoms associated to victims of sexual abuse i.e., depressive and anxiety disorders. Nevertheless, the race or ethnicity of victims of sexual abuse was not related in itself to the manifestation of depressive or anxiety disorders (Mennen, 1995).

In the initial search for studies in the meta-search engines, both English and Spanish descriptors were used: child maltreatment/maltrato infantil, child sexual abuse/abuso sexual infantil, victimization/victimización, internalizing disorders/trastornos internalizantes, anxiety/ansiedad and depression/depresión. In the second search in the Web of Science, the same English language descriptors were used given that all of the descriptors in these databases are in English. Following a method of successive approximations, all of the keywords were reviewed in the selected articles in the search for other potential descriptors. This method identified other descriptors employed by other authors (e.g., child sexual abuse, adolescent sexual abuse, internalizing behaviour disorder, CSA survivors, and sequelae) that were included in the search. In all, the system yielded more than 15,900 (searching by child sexual abuse) and 2,000 studies (searching by child sexual abuse AND depression OR anxiety) that were finally reduced to 78 after applying the following inclusion and exclusion criteria.

### Inclusion and exclusion criteria

Of the studies listed by the system, the following met the inclusion criteria: a) studies assessing the sequelae of CSA/ASA [CSA/ASA understood as the involvement of a child in sexual activity that he or she does not fully comprehend, is unable to give informed consent to, or for which the child is not developmentally prepared and cannot give consent] in terms of depression or anxiety internalizing variables; b) studies reporting the effect sizes of CSA/ASA sequelae, in which the variables and/or statistics enabled the following to be

calculated: group size, mean and standard deviation of sequelae measurement variables for each group, prevalence, specificity, and sensitivity; c) studies defining the ground truth for classifying participants as victims of CSA/ASA or the measure (i.e., instrument) of abuse; and d) studies providing descriptive data on the sample employed (e.g., age, sample size).

Studies failing to meet the prescribed requirements were excluded, as were cases where, after contacting the authors, the data required was not facilitated for the computation of the effect sizes. In addition, studies with data errors (e.g., lack of consistency in group size throughout the study not attributable to missing data) were eliminated. Similarly, studies failing to guarantee the mutual exclusion of the victim of sexual abuse condition from other forms of maltreatment were also excluded (e.g., studies undertaking a single comprehensive analysis of victims of sexual and physical abuse or neglect). The inclusion of studies was restricted in time to studies published since 1995 given the profusion of meta-analysis on studies up to 1995 (Jumper, 1995). Thus, by applying these criteria, 78 studies were selected, with a total of 19,360 subjects, from which 149 effect sizes were obtained: 62 for the effects of CSA/ASA on anxiety disorders, and 87 for depressive disorders. All of the studies selected had been published in double blind peer-reviewed scientific journals indexed at the Web of Science.

### **Coding of primary studies**

The following data from the studies was coded for the meta-analysis: variables measuring the effects of abuse (i.e., clinical diagnosis, clinical symptoms), measures of abuse, reliability of the measurement instruments, sample characteristics (i.e., size, age, cultural context), and the statistics required for computing the effect sizes.

All of the studies of the sample were examined by two independent researchers, with total agreement in their classifications ( $k = 1$ ). Appendix 1 shows the characteristics of the primary studies included in the present meta-analysis.

### **Data analysis**

The weighted effect size was calculated according to the sample size of the variable measuring the effects of CSA/ASA (anxiety or depression). When several measures of anxiety were reported (e.g., generalized, social phobia, specific phobia) or depression (e.g., major depression, dysthymia), the means for the effect sizes of the variables with more than one measure were calculated. Moreover, when the sample was subdivided into subtypes of abuse, the effect sizes were weighted in order to obtain an overall size.

The meta-analytical technique employed was a correlational type procedure (Hunter and Schmidt, 2004), weighting the effect size with the sample size and correcting with the reliability of the predictor and the criterion.

As most of studies fail to provide correlations between sexual abuse and the measures of sequelae of internalizing disorders (i.e., anxiety and depression), an alternative estimator of effect size i.e., Cohen's  $d$  (1988), was computed when studies reported the mean and standard deviations of the experimental group (abused children), but no control group (in these studies the control group was taken to be the normative population for each instrument). When the

results were expressed as student  $t$  values ( $F$  for the degree of freedom was transformed into  $t$  values), Cohen's  $d$  was obtained by comparing groups of the same sizes or when homogeneity of variance was observed and Hedges'  $g$  (Hedges & Olkin, 1985) were obtained for different sizes or no homogeneity of variance. If the results were expressed in  $Z$  values, they were transformed into correlations; phi was obtained from 2X2 tables; and when in Odds Ratio, converted to correlations. The formulas for converting the other effect sizes to correlations were taken from Cohen (1988) and Rosenthal (1994).

To estimate the practical utility of the results, the  $U1$  (Cohen, 1988), the Odds Ratio, and the CLES (McGraw & Wong, 1992) were calculated. Additionally, increases in CSA/ASA symptoms or disorders (injury quantified in relation to population baselines) were directly obtained from  $r$ .

### Predictor reliability

For predictor reliability, the measure of sexual abuse, the reliability coefficients obtained from the primary studies were computed. When studies failed to report reliability coefficients or reported concordance, which is not reliability, mean reliability was calculated on the basis of the primary studies. Table 1 shows mean reliability, the standard error, and a 95% confidence interval (indicating the variance of mean reliability) for each context measured. Succinctly, the measure of the predictor, victims of sexual abuse, was reliable (see Table 1) with a limit below .77 (that is, high reliability,  $r > .70$ ; Arce, Velasco, Novo, & Fariña, 2014; Nunnally & Bernstein, 1994).

Table 1. Predictor Reliability

Measure	$r_{xx}$	$SEM_{r_{xx}}$	95% IC	$n$
Sexual Abuse Victimization	.82	.023	.77, .86	17

Note.  $r_{xx}$  = average measure reliability;  $SEM_{r_{xx}}$  = mean standard error; 95% CI = 95% confidence interval;  $n$  = number of reliability coefficients.

### Criterion reliability

Criterion reliability was drawn from the primary studies, the original publication of the instrument itself and, in the absence of both, completed with the means for those contingencies where they were unavailable. The mean reliability, standard error, and confidence intervals for each measure are shown in Table 2. In short, the measure of criteria reliability ranged from .84 to .86, with a lower limit for a 95% confidence interval of .82. Thus the measures of the criterion were highly reliable ( $r > .70$ ).

Table 2. Criterion Reliability

Criteria	$r_{yy}$	$SEM_{r_{yy}}$	95% CI	$n$
Anxiety	.84	.010	.82, .86	41
Depression	.86	.008	.84, .88	57
Total	.86	.006	.85, .87	61

Note.  $r_{yy}$  = average measure reliability;  $SEM_{r_{yy}}$  = mean standard error; 95% CI = 95% confidence interval;  $n$  = number of reliability coefficients.

## Results

### Study of outliers

Outliers in each of the measures (general sequelae, depression, and anxiety) were removed using the criterion  $\pm 2 SD$  (bilateral) of the mean effect size so the results were generalizable to 96% of future samples. The results identified 6 outliers ( $-.12 > r > .64$ ) in sequelae ( $M = .26, SD = .19$ ); 5 outliers ( $-.13 > r > .63$ ) in anxiety ( $M = .25, SD = .19$ ); and 6 outliers ( $-.14 > r > .66$ ) in depression ( $M = .26, SD = .20$ ). Thus, a total of 6 studies were eliminated from the meta-analysis as outliers.

### Global analysis

The results of the meta-analysis on sequelae of sexual abuse on mental health (internalizing symptoms: depression and anxiety) showed (see Table 3) a significant effect (when the confidence interval has no zero, indicating the effect size was significant), positive (between victimization and mental health injury), generalizable (when the credibility interval has no zero, indicating the effect size was generalizable to 90% of other samples), and of a medium size (Cohen's category:  $r = .30$ ) in general sequelae, depression, and anxiety, explaining 12, 8, and 10% of the variance, respectively. In terms of practical utility, victims of CSA/ASA had a 70% higher probability (CLES) of internalizing injury (general sequelae) than non-victims; with 66% for depression and 68% for anxiety. The injury caused was quantified as 34, 28 and 31% in general sequelae, depression, and anxiety, respectively. Finally, the distributions for victims and non-victims (U1) in general sequelae, depression, and anxiety, were totally independent (44, 37 and 41%, respectively), though they were expected to be similar.

As for the robustness of the significance of the sequelae in CSA/ASA victims, it is worth noting that 93 studies with no significant results would be required to accept the null hypothesis.

Though results were generalizable in three measures (general sequelae, depression, and anxiety), the literature recommends assessing the potential differential effects of gender, type of measure, and type of abuse (meta-analysis). In short, it is well known that the base rate for symptoms and clinical diagnosis differ from males to females (Nolen-Hoeksema, 1990, 2002). This view is so firmly established that most of the psychometric measurement instruments extensively used in clinical practice score males separately to females. Likewise, according to the leading international organisations classifying mental disorders (American Psychiatric Association, APA, 2013; OMS, 2000), there is a higher prevalence of diagnosed depression and anxiety among females than males, including school-aged children and adolescents. Similarly, the type of abuse suffered (i.e., non-contact, contact, intercourse) has been shown to have effects on sequelae, which relates the severity of the crime to the severity of the injury (mild, moderate, severe) (Bulik et al., 2001; Cutajar et al., 2010; Fergusson et al., 2013; Jonas et al., 2011). As for the variables measuring the effects on victims of abuse i.e., clinical diagnosis and clinical symptoms, the presence of symptoms does not imply that the diagnostic criteria had been met. In other words, they are different measures given that they measure different constructs, so they may have a differentially sensitive to injury.

## Study of moderators

### Gender effects

The results of the meta-analysis showed (see Table 4) a significant effect, positive, of a small size (Cohen's category:  $r = .10$ ), and generalizable in depression and anxiety in female CSA/ASA victims.

In comparison, the meta-analysis revealed for male CSA/ASA victims a significant effect, positive, and of a small size in depression and anxiety, being generalizable in depression, but not so in anxiety (see Table 4). Thus, in the latter case, the results exhibited moderators mediated the direction of the effects.

Having contrasted the significance of the differences between the effect sizes, the true correlation between female participants and male participants, sequelae in depression was found to be significantly higher,  $q_s = 0.093$ ,  $p < .05$ , in females, but not so for anxiety,  $q_s = 0.073$ , *ns*. This translates into quantifying injury in females as suffering from 9% more injury in depression than males. In prevalence rates (odds ratio), injury in depression for female and male victims was 2.26 and 1.60 times greater than for non-victims, and 2.26 and 1.73 times greater in anxiety for female and male victims, respectively, in contrast to non-victims.

### Effects of the type of measure

The results of the meta-analysis showed a significant effect, positive, of a medium size, and generalizable in the diagnosis of depressive and anxiety disorders in victims of CSA/ASA (see Table 5).

Likewise, the results of the meta-analysis displayed a significant effect, positive, of a small effect size and generalizable in anxiety and depression symptoms in CSA/ASA victims (see Table 5).

Comparatively, injury was significantly higher,  $q_s = 0.152$ ,  $p < .01$ , in the clinical diagnosis of anxiety disorder than reported symptoms of anxiety. Likewise, the diagnosis of a depressive disorder was significantly more sensitive,  $q_s = 0.108$ ,  $p < .05$ , for CSA/ASA victims than the report of depressive symptoms. These results, due to the type of measure, explained the differences attributed to sample type (Rind, Bauserman, & Tromovitch, 1998): university population (measure of symptoms), and clinical population (clinical diagnosis).

The meta-analysis of major depressive disorder and dysthymia (persistent depressive disorder) nesting in the diagnosis of depression (see Table 5), confirmed a significant and positive effect, of a medium size, and generalizable. Thus, the prevalence of a dysthymic disorder in victims of abuse was 6.59 (odds ratio) times higher than for non-victims, 3.25 higher for major depression. In terms of injury quantification, it was of 46% for dysthymia, and 31% for major depression. The difference between effect sizes was significant,  $q_s = 0.176$ ,  $p < .01$ , thus the effect size for the diagnosis of dysthymia was significantly larger than for major depressive disorder.



Similarly, the meta-analysis on generalized anxiety disorder, specific phobia, social phobia and panic disorder were nested in anxiety disorders (see Table 5) found a significant and positive effect, of a medium to large, and generalizable for every diagnosis. Indeed, this implied CSA/ASA victims had 5.12, 7.62, 4.85, and 5.60 (odds ratio) greater probability of developing generalized anxiety disorder, specific phobia, social phobia, and panic disorder, respectively, than CSA/ASA non-victims. Injury was quantified as 41, 49, 40, and 43% for generalized anxiety disorder, specific phobia, social phobia and panic disorder, respectively. The probability of developing these disorders as sequelae was similar except for specific phobia that was significantly higher than social phobia,  $q_s = 0.112$ ,  $p < .05$ , and generalized anxiety disorder,  $q_s = 0.100$ ,  $p < .05$ .

### **Effects of the type of abuse**

The results of the meta-analysis on the type of abuse suffered (no contact, contact, and penetration) revealed a significant and positive effect, of small size, and generalizable in depression and anxiety (see Table 6). The comparison of sizes, showed injury derived from abuse with penetration, both in depression and anxiety, was significantly higher than injury in the no contact abuse condition for depression,  $q_s = 0.093$ ,  $p < .05$ ; and anxiety,  $q_s = 0.092$ ,  $p < .05$ .

### **Effects of the interaction type of measure and gender**

In the diagnosis of depressive disorders (see Table 7), the effect sizes were positive and significant for both males and females, of a small size for males and a medium one for females, which were generalizable for females, but not for males (the effects of the moderators could not be assessed in this case due to the very small  $k$ ). In comparison, the effect size for females was significantly higher,  $q_s = 0.388$ ,  $p < .01$ , than for males, quantifying injury in 42% for female victims and 10% for males. As for prevalence, female CSA/ASA victims had a 5.40 (odds ratio) more probability of meeting the criteria of depressive disorders than female non-victims, whereas males had a 1.44 more probability than male non-victims. Moreover, for the diagnosis of anxiety disorders the effect sizes were positive, significant, and of a small size for both males and females, generalizable for females, but not so for males (once again, moderators could not be found due to the very small  $k$ ). Once again, the effect size found in females was significantly higher,  $q_s = 0.104$ ,  $p < .05$ , than in males, with injury of 24% for female victims and 14% for males. This reveals that female victims had 2.43 (odds ratio) more probability of developing anxiety disorders than female non-victims, and male victims 1.66 more probability than male non-victims.

The effect sizes in depressive symptoms were significant, positive, of small sizes, generalizable, and similar (*ns*) for both males and females. As for anxiety symptoms, the effect sizes were significant, positive, of small sizes, and generalizable for both males and females. Nonetheless, the effect size was significantly higher,  $q_s = 0.095$ ,  $p < .05$ , in males. Thus, the results highlight that male CSA/ASA victims had a 2.76 probability (odds ratio) of developing significantly more anxiety symptoms than male non-victims, and female victims a 1.95 probability than female non-victims.

**Table 3. Results of the Meta-Analyses of Sexual Abuse Victimization in General Sequelae, Depression and Anxiety**

	k	N <sub>E</sub>	N <sub>C</sub>	N <sub>T</sub>	r <sub>w</sub>	SD <sub>r</sub>	ρ	SD <sub>p</sub>	%VE	95% CI <sub>r</sub>	90% CI <sub>p</sub>
General Sequelae	91	19360	93988	125555	.28	.17	.34	.20	2.82	[.27, .29]	[.08, .59]
Depression	87	18910	92618	123735	.24	.14	.28	.16	4.10	[.23, .25]	[.08, .49]
Anxiety	62	14587	77494	93075	.26	.14	.31	.17	3.73	[.25, .27]	[.10, .52]

Note. k = number of studies; N<sub>E</sub> = experimental group sample size; N<sub>C</sub> = control group sample size; N<sub>T</sub> = total sample size; r<sub>w</sub> = observed correlation (observed validity) weighted for sample size; SD<sub>r</sub> = standard deviation of the observed correlation; ρ = true correlation (operational validity corrected for criterion and predictor unreliability); SD<sub>p</sub> = standard deviation of true correlation; %VE = percentage of variance accounted for by artifactual errors; 95% CI<sub>r</sub> = 95% confidence interval; 90% CI<sub>p</sub> = 90% credibility interval. When N<sub>T</sub> ≠ N<sub>E</sub> + N<sub>C</sub>, it means that experimental or control group sample size in primary studies were unknown.

**Table 4. Results of the Meta-Analyses of Sexual Abuse Victimization in Depression and Anxiety by Gender**

	k	N <sub>E</sub>	N <sub>C</sub>	N <sub>T</sub>	r <sub>w</sub>	SD <sub>r</sub>	ρ	SD <sub>p</sub>	%VE	95% CI <sub>r</sub>	90% CI <sub>p</sub>
Depression Measure											
Females	42	8074	20127	39498	.18	.09	.22	.09	14.56	[.17, .19]	[.10, .34]
Males	12	1830	13843	15673	.11	.08	.13	.10	10.60	[.09, .13]	[.01, .26]
Anxiety Measure											
Females	27	4926	12542	17706	.18	.12	.22	.13	10.82	[.17, .19]	[.05, .39]
Males	8	998	7380	8378	.12	.13	.15	.15	5.56	[.09, .14]	[-.04, .35]

Note. k = number of studies; N<sub>E</sub> = experimental group sample size; N<sub>C</sub> = control group sample size; N<sub>T</sub> = total sample size; r<sub>w</sub> = observed correlation (observed validity) weighted for sample size; SD<sub>r</sub> = standard deviation of the observed correlation; ρ = true correlation (operational validity corrected for criterion and predictor unreliability); SD<sub>p</sub> = standard deviation of true correlation; %VE = percentage of variance accounted for by artifactual errors; 95% CI<sub>r</sub> = 95% confidence interval; 90% CI<sub>p</sub> = 90% credibility interval. When N<sub>T</sub> ≠ N<sub>E</sub> + N<sub>C</sub>, it means that experimental or control group sample size in primary studies were unknown.

**Table 5. Results of the Meta-Analyses of Sexual Abuse Victimization in Depression and Anxiety by Type of Measure.**

	k	N <sub>E</sub>	N <sub>C</sub>	N <sub>T</sub>	r <sub>w</sub>	SD <sub>r</sub>	ρ	SD <sub>p</sub>	%VE	95% CI <sub>r</sub>	90% CI <sub>p</sub>
Depressive Disorder	28	12131	66986	90220	.26	.14	.31	.03	2.32	[.25, .27]	[.10, .52]
Dysthymia	8	4524	36668	41192	.38	.08	.46	.09	8.91	[.37, .39]	[.34, .56]
Major Depressive Disorder	24	9406	64284	84793	.26	.14	.31	.16	2.21	[.25, .27]	[.11, .52]
Depressive symptomatology	59	6668	25533	33293	.18	.12	.21	.13	12.75	[.17, .19]	[.04, .37]
Anxiety Disorder	21	10133	58784	68917	.29	.14	.35	.16	2.40	[.28, .30]	[.14, .56]
Generalized Anxiety Disorder	8	5808	43403	49211	.34	.11	.41	.13	3.63	[.33, .35]	[.25, .57]
Specific Phobia	3	3830	30616	34446	.41	.03	.49	.02	70.97	[.40, .42]	[.46, .51]
Social Phobia	10	4901	39701	44602	.34	.13	.40	.15	2.82	[.33, .35]	[.21, .60]
Panic Disorder	8	4932	37321	42253	.36	.11	.43	.12	4.30	[.35, .37]	[.27, .58]
Anxiety symptomatology	41	4510	18845	24270	.18	.12	.21	.13	12.53	[.17, .19]	[.05, .37]

Note. k = number of studies; N<sub>E</sub> = experimental group sample size; N<sub>C</sub> = control group sample size; N<sub>T</sub> = total sample size; r<sub>w</sub> = observed correlation (observed validity) weighted for sample size; SD<sub>r</sub> = standard deviation of the observed correlation; ρ = true correlation (operational validity corrected for criterion and predictor unreliability); SD<sub>p</sub> = standard deviation of true correlation; %VE = percentage of variance accounted for by artifactual errors; 95% CI<sub>r</sub> = 95% confidence interval; 90% CI<sub>p</sub> = 90% credibility interval. When N<sub>T</sub> ≠ N<sub>E</sub> + N<sub>C</sub>, it means that experimental or control group sample size in primary studies were unknown. Effect size for agoraphobia has not been obtained due to insufficient k.

**Table 6. Results of the Meta-analyses of Sexual Abuse Victimization in Depression and Anxiety by Type of Abuse.**

	k	N <sub>E</sub>	N <sub>C</sub>	N <sub>T</sub>	r <sub>w</sub>	SD <sub>r</sub>	ρ	SD <sub>p</sub>	%VE	95% CI <sub>r</sub>	90% CI <sub>p</sub>
Depression Measure											
Non-Contact	7	278	5431	5709	.12	.03	.14	0	100	[.09, .15]	[.14]
Contact	4	171	3228	3399	.16	.07	.18	.07	25.97	[.13, .19]	[.10, .27]
Intercourse	4	184	3228	3412	.19	.08	.23	.09	17.60	[.16, .22]	[.11, .34]
Anxiety Measure											
Non-Contact	4	101	3225	3326	.08	.03	.09	0	100	[.05, .11]	[.09]
Contact	4	170	3225	3395	.11	.06	.14	.06	34.73	[.08, .14]	[.06, .21]
Intercourse	4	184	3225	3409	.15	.07	.18	.08	22.02	[.12, .18]	[.08, .28]

Note. k = number of studies; N<sub>E</sub> = experimental group sample size; N<sub>C</sub> = control group sample size; N<sub>T</sub> = total sample size; r<sub>w</sub> = observed correlation (observed validity) weighted for sample size; SD<sub>r</sub> = standard deviation of the observed correlation; ρ = true correlation (operational validity corrected for criterion and predictor unreliability); SD<sub>p</sub> = standard deviation of true correlation; %VE = percentage of variance accounted for by artifactual errors; 95% CI<sub>r</sub> = 95% confidence interval; 90% CI<sub>p</sub> = 90% credibility interval. When N<sub>T</sub> ≠ N<sub>E</sub> + N<sub>C</sub>, it means that experimental or control group sample size in primary studies were unknown.

**Table 7. Results of the Meta-analyses of the Sexual Abuse Victimization in Depression and Anxiety by Type of Measure and Gender.**

	k	N <sub>E</sub>	N <sub>C</sub>	N <sub>T</sub>	r <sub>w</sub>	SD <sub>r</sub>	p	SD <sub>p</sub>	%VE	95% CI <sub>r</sub>	90% CI <sub>p</sub>
<b>Depressive Disorder Diagnosis</b>											
Females	11	4421	11594	27130	.33	.19	.42	.23	1.79	[.32, .34]	[.13, .71]
Males	5	1446	10220	11666	.08	.08	.10	.09	6.90	[.06, .10]	[-.02, .21]
<b>Anxiety Disorder Diagnosis</b>											
Females	8	3192	8018	11210	.20	.13	.24	.15	4.75	[.18, .22]	[.05, .43]
Males	4	845	6806	7651	.11	.13	.14	.15	3.26	[.09, .13]	[-.06, .33]
<b>Depressive Symptomatology</b>											
Females	32	3709	8530	12482	.17	.09	.20	.09	27.79	[.15, .19]	[.07, .32]
Males	7	458	3577	4035	.18	.05	.22	.04	62.72	[.15, .20]	[.17, .26]
<b>Anxiety Symptomatology</b>											
Females	20	1790	4580	6608	.15	.09	.18	.09	33.68	[.13, .17]	[.07, .30]
Males	4	153	594	727	.23	.09	.27	.05	69.51	[.16, .30]	[.20, .35]

Note. k = number of studies; N<sub>E</sub> = experimental group sample size; N<sub>C</sub> = control group sample size; N<sub>T</sub> = total sample size; r<sub>w</sub> = observed correlation (observed validity) weighted for sample size; SD<sub>r</sub> = standard deviation of the observed correlation; p = true correlation (operational validity corrected for criterion and predictor unreliability); SD<sub>p</sub> = standard deviation of true correlation; %VE = percentage of variance accounted for by artifactual errors; 95% CI<sub>r</sub> = 95% confidence interval; 90% CI<sub>p</sub> = 90% credibility interval. When N<sub>T</sub> ≠ N<sub>E</sub> + N<sub>C</sub>, it means that experimental or control group sample size in primary studies were unknown.

## Discussion

As a whole, the results of this study support undoubtedly (a total of 93 studies with non-significant results would be required to annul the effect) that CSA/ASA victimization had a significant and positive effect (injury) on mental health, of a small- to large-size, and generalizable. This was demonstrated in the following:

- a) A higher probability, around 70% in each of the different measures, of suffering from internalized injury, depression, and anxiety.
- b) Injury caused to the victim's mental health, that is, mental injury and/or emotional suffering (United Nations, 1988), was calculated to be around 30% (34, 28 and 31% in general sequelae, depression, and anxiety, respectively). This finding implies that offenders are not only criminally responsible for their deeds, but are also liable to civil compensation payments for injuries caused to victims. With this aim in mind a forensic technique has been developed for quantifying injury in specific cases (Arce & Fariña, 2009).
- c) Injury to mental health in terms of depression and anxiety associated to victims of CSA/ASA was significant in males and females, but with 9% more depression in female, leading to a higher probability of developing a depressive or anxiety disorders in females with injury of 42 and 24%, respectively. In contrast, injury involved significantly more anxiety symptoms in males with 27% injury. However, symptoms are not an optimum indicator of injury.
- d) Clinical diagnosis was a measure of injury significantly more adequate than symptoms. The evaluation techniques characteristic to clinical diagnosis and clinical symptoms may explain these differences. In the interview, indeed, injury was linked to cause; whereas in the psychometric measures of CSA/ASA victimization it was not, allowing for other causes. Furthermore, the diagnostic threshold was much stricter than for symptoms, which underscores its greater sensitivity and specificity. Thus, the benchmark for future research should be the diagnosed measure of injury based on an interview task, rather than symptoms based on a psychometric measure.
- e) Injury was calculated to be 46 and 31% for persistent depressive disorder (dysthymia) and major depressive disorder, respectively. Moreover, the expression of injury as dysthymia was significantly greater than for major depressive disorder, that is, the probability of chronic injury (dysthymic) was greater than more serious injury (major depressive disorder).
- f) Injury to CSA/ASA victims was expressed in anxiety disorders, estimated to be around 40 to 49%, being the highest in specific phobias.
- g) Abuse with penetration led to injury in depression and anxiety significantly greater than abuse with no contact. These results lend support to the distinction in the legal classification of both criminal typologies.

### Limitations of the study

This meta-analysis entails certain limitations that should be borne in mind when interpreting the data. First, the ground truth of the primary studies for the classification of abuse generally rests on self-reports of a retrospective nature that relies on individual memory capabilities, and are related to false positives or false alarms (Amado, Arce, & Fariña. 2015;

Schoedl et al., 2010). Moreover, victim self-reports of sexual abuse may bias the results towards concealing them (false negatives), in particular for males (Stoltenborgh et al., 2011). Second, primary studies assume that injury to mental health is sequelae to abuse, without appraising other possible causes (cause-effect relationship) (Jumper, 1995; Vilariño et al., 2013). Third, the effect of the variable under analysis in primary studies was not completely isolated as in many studies victims of sexual abuse, physical abuse, neglect, and other categories appear under the same umbrella. Fourth, as some studies had no control group, the normative population was taken as the contrast group; or it was not equivalent to the experimental one with the subsequent potential for distortion in the calculated effect sizes (Briere, 1992).

Alternatively, the results of the meta-analysis were subject to little variability, that is,  $N_s > 400$  and a large  $k$  (Hunter & Schmidt, 2004), were highly generalizable (entirely for the female population, and for males with the exception of the diagnosis of a disorder and the general measure of anxiety for the male population), whereas 93 studies with no significant results would be required to annul the evidence supporting the claim that CSA/ASA leads to mental health injuries.

Further research is required to determine which moderators inhibit the generalization of the effects in the general measure of anxiety in the male population and in the diagnosis of depression and anxiety.

## References

[References marked with an asterisk indicate studies included in the meta-analysis]

- Amado, B. G., Arce, R., & Fariña (2015). Undeutsch hypothesis and Criteria Based Content Analysis: A meta-analytic review. *The European Journal of Psychology Applied to Legal Context*, 7, 3-12. doi: 10.1016/j.ejpal.2014.11.002
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: American Psychiatric Association.
- Arce, R., & Fariña, F. (2009). Evaluación psicológica forense de la credibilidad y daño psíquico en casos de violencia de género mediante el Sistema de Evaluación Global. In F. Fariña, R. Arce, & G. Buela-Casal (Eds.), *Violencia de género. Tratado psicológico y legal* (pp. 147-168). Madrid: Biblioteca Nueva.
- Arce, R., Velasco, J., Novo, M., & Farina, F. (2014). Elaboración y validación de una escala para la evaluación del acoso escolar [Development and validation of a scale to assess bullying]. *Revista Iberoamericana de Psicología y Salud*, 5, 71-104.
- \*Balsam, K. F., Lehavot, K., Beadnell, B., & Circo, E. (2010). Childhood abuse and mental health indicators among ethnically diverse lesbian, gay, and bisexual adults. *Journal of Consulting and Clinical Psychology*, 74, 459-468. doi: 10.1037/a0018661
- \*Bonomi, A. E., Cannon, E. A., Anderson, M. L., Rivara, F. P., & Thompson, R. S. (2008). Association between self-reported health and physical and/or sexual abuse experienced before age 18. *Child Abuse & Neglect*, 32, 693-701. doi: 10.1016/j.chiabu.2007.10.004
- Briere, J. (1992). Methodological issues in the study of sexual abuse effects. *Journal of Consulting and Clinical Psychology*, 60, 196-203. doi: 10.1037/0022-006X.60.2.196

- \*Briere, J., & Elliott, D. M. (2003). Prevalence and psychological sequelae of self-reported childhood physical and sexual abuse in a general population sample of men and women. *Child Abuse & Neglect*, 27, 1025-1222. doi: 10.1016/j.chiabu.2003.09.008
- \*Brown, J., Cohen, P., Johnson, F., & Smailes, E. (1999). Childhood abuse and neglect: Specificity of effects on adolescent and young adult depression and suicidality. *Journal of the American of child & Adolescent Psychiatry*, 38, 1490-1496. doi: 10.1097/00004583-199912000-00009
- Bulik, C. M., Prescott, C. A., & Kendler, K. S. (2001). Features of childhood sexual abuse and the development of psychiatric and substance use disorders. *British Journal of Psychiatry*, 179, 444-449. doi: 10.1192/bjp.179.5.444
- \*Cantón-Cortés, D., Cortés, M. R., & Cantón, J. (2012). The role of traumagenic dynamics on the psychological adjustment of survivors of child sexual abuse. *European Journal of Developmental Psychology*, 9, 665-680. doi: 10.1080/17405629.2012.660789
- \*Cantón-Cortés, D., & Justicia-Justicia, F. (2008). Afrontamiento del abuso sexual infantil y ajuste psicológico a largo plazo [Child sexual abuse coping and long term psychological adjustment]. *Psicothema*, 20, 209-515.
- \*Carey, P. D., Walker, J. L., Rossouw, W., Seedat, S., Stein, D. J. (2008). Risk indicators and psychopathology in traumatised children and adolescents with a history of sexual abuse. *European Child & Adolescent Psychiatry*, 17, 93-98. doi: 10.1007/s00787-007-0641-0
- \*Cheasty, M., Clare, A. W., & Collins, C. (1998). Relation between sexual abuse in childhood and adult depression: Case-control study. *British Medical Journal*, 316, 198-201. doi: 10.1136/bmj.316.7126.198
- \*Chen, J., Cai, Y. Y., Cong, E., Liu, Y., Gao, J., Li, Y.,...Flint, J. (2014). Childhood sexual abuse and the development of recurrent major depression in Chinese women. *Plos One*, 9(1), e87569. doi: 10.1371/journal.pone.0087569
- \*Chen, J., Dunne, M. P., & Han, P. (2004). Child sexual abuse in China: A study of adolescents in four provinces. *Child Abuse & Neglect*, 28, 1171-1186. doi:10.1016/j.chiabu.2004.07.003
- \*Chen, J., Dunne, M. P., & Han, P. (2006). Child sexual abuse in Henan province, China: Associations with sadness, suicidality, and risk behaviors among adolescent girls. *Journal of Adolescent Health*, 38, 544-549. doi: 10.1016/j.jadohealth.2005.04.001
- Cohen, J. (1988). *Statistical power analysis for behavioral sciences* (2nd ed.). New York, NY: Academic Press.
- \*Comijs, H. C., van Exel, E., van der Mast, R. C., Paauw, A., Voshaar, R. O., & Stek, M. L. (2013). Childhood abuse in late-life depression. *Journal of Affective Disorders*, 147(1-3), 241-246. doi: 10.1016/j.jad.2012.11.010
- \*Cortés-Arboleda, M. R., Cantón-Cortés, D., & Cantón-Duarte, J. (2011). Consecuencias a largo plazo del abuso sexual infantil: Papel de la naturaleza y continuidad del abuso y del ambiente familiar [Long term consequences of child sexual abuse: The role of the nature and continuity of abuse and family environment]. *Behavioral Psychology-Psicología Conductual*, 19, 41-56.
- \*Cortés-Arboleda, M. R., Cantón-Duarte, J., & Cantón-Cortés, D. (2011). Naturaleza de los abusos sexuales a menores y consecuencias en la salud mental de las víctimas [Characteristics of sexual abuse of minors and its consequences on victims' mental health]. *Gaceta Sanitaria*, 25, 157-165. doi: 10.1016/j.gaceta.2010.10.009
- \*Cutajar, M. C., Mullen, P. E., Ogloff, J. R. P., Thomas, S. D., Wells, D. L., & Spataro, J. (2010). Psychopathology in a large cohort of sexually abused children followed up to 43 years. *Child Abuse & Neglect*, 34, 813-822. doi: 10.1016/j.chiabu.2010.04.004

- \*Doerfler, L. A., Toscano Jr., P. F., & Connor, D. F. (2009). Sex and aggression: The relationship between gender and abuse experience in youngsters referred to residential treatment. *Journal of Child and Family Studies, 18*, 112-122. doi: 10.1007/s10826-008-9212-3
- \*Dube, S. R., Anda, R. F., Whitfield, C. L., Brown, D. W., Felitti, V. J., Dong, M., & Giles, W. H. (2005). Long-term consequences of childhood sexual abuse by gender of victim. *American Journal of Preventive Medicine, 28*, 430-438. doi: 10.1016/j.amepre.2005.01.015
- \*Feeney, J., Kamiya, Y., Robertson, I. H., & Kenny, R. A. (2013). Cognitive function is preserved in older adults with a reported history of childhood sexual abuse. *Journal of Traumatic Stress, 26*, 735-743. doi 10.1002/jts.21861
- \*Feerick, M. M., & Snow, K. L. (2005). The relationships between childhood sexual abuse, social anxiety, and symptoms of posttraumatic stress disorder in women. *Journal of Family Violence, 20*, 409-419. doi: 10.1007/s10896-005-7802-z
- \*Ferguson, K. S., & Dacey, C. M. (1997). Anxiety, depression, and dissociation in women health care providers reporting a history of childhood psychological abuse. *Child Abuse & Neglect, 21*, 941-952. doi: 10.1016/S0145-2134(97)00055-0
- \*Fergusson, D. M., Boden, J. M., & Horwood, L. J. (2008). Exposure to childhood sexual and physical abuse and adjustment in early adulthood. *Child Abuse & Neglect, 32*, 607-619. doi:10.1016/j.chiabu.2006.12.018
- \*Fergusson, D. M., McLeod, G. F. H., & Horwood, L. J. (2013). Childhood sexual abuse and adult developmental outcomes: Findings from a 30-year longitudinal study in New Zealand. *Child Abuse & Neglect, 37*, 664-674. doi 10.1016/j.chiabu.2013.03.013
- \*Fondacaro, K. M., Holt, J. C., & Powell, T. A. (1999). Psychological impact of childhood sexual abuse on male inmates: The importance of perception. *Child Abuse & Neglect, 23*, 361-369. doi: 10.1016/S0145-2134(99)00004-6
- \*Frías, M. T., Brassard, A., & Shaver, P. R. (2014). Childhood sexual abuse and attachment insecurities as predictors of women's own and perceived-partner extradyadic involvement. *Child Abuse & Neglect, 38*, 1450-1458. doi: 10.1016/j.chiabu.2014.02.009
- \*Godbout, N., Briere, J., Sabourin, S., & Lussier, Y. (2014). Child sexual abuse and subsequent relational and personal functioning: The role of parental support. *Child Abuse & Neglect, 38*, 317-325. doi: 10.1016/j.chiabu.2013.10.001
- \*Gudjonsson, G. H., Sigurdsson, J. F., & Tryggvadóttir, H. B. (2011). The relationship of compliance with a background of childhood neglect and physical and sexual abuse. *The Journal of Forensic Psychiatry & Psychology, 22*, 87-98. doi: 10.1080/14789949.2010.524707
- \*Haj-Yahia, M. M., & Tamish, S. (2001). The rates of child sexual abuse and its psychological consequences as revealed by a study among Palestinian university students. *Child Abuse & Neglect, 25*, 1303-1327. doi: 10.1016/S0145-2134(01)00277-0
- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.
- \*Henderson, D., Hargreaves, I., Gregory, S., & Williams, J. M. G. (2002). Autobiographical memory and emotion in a non-clinical sample of women with and without a reported history of childhood sexual abuse. *British Journal of Clinical Psychology, 41*, 129-141. doi: 10.1348/014466502163921
- \*Hobfoll, S. E., Bansal, A., Schurg, R., Young, S., Pierce, C. A., Hobfoll, I., & Johnson, R. (2002). The impact of perceived child physical and sexual abuse history on native



- American women's psychological well-being and AIDS risk. *Journal of Consulting and Clinical Psychology*, 70, 252-257. doi: 10.1037//0022-006X.70.1.252
- Hunter, J. E., & Schmidt, F. L. (2004). *Methods of Meta-analysis: Correcting error and bias in research findings* (2nd ed.). Thousand Oaks, CA: Sage.
- Intebi, I. V. (1998). *Abuso sexual infantil: En las mejores familias*. Buenos Aires, Argentina: Granica.
- \*Jonas, S., Bebbington, P., McManus, S., Meltzer, H., Jenkins, R., Kuipers, E., Cooper, C., King, M., & Brugha, T. (2011). Sexual abuse and psychiatric disorder in England: Results from the 2007 adult psychiatric morbidity survey. *Psychological Medicine*, 41, 709-719. doi: 10.1017/S003329171000111X
- Jumper, S. A. (1995). A meta-analysis of the relationship of child sexual abuse to adult psychological adjustment. *Child Abuse & Neglect*, 19, 715-728. doi: 10.1016/0145-2134(95)00029-8
- \*Kendler, K. S., Bulik, C. M., Silberg, J., Hettema, J. M., Myers, J., & Prescott, C. A. (2000). Childhood sexual abuse and adult psychiatric and substance use disorders in women. An epidemiological and cotwin control analysis. *Archives of General Psychiatry*, 57, 953-959. doi: 10.1001/archpsyc.57.10.953
- \*Kent, A., & Waller, G. (1998). The impact of childhood emotional abuse: An extension of the child abuse and trauma scale. *Child Abuse & Neglect*, 22, 393-399. doi: 10.1016/S0145-2134(98)00007-6
- Koenen, K. C., & Widom, C. S. (2009). A prospective study of sex differences in the lifetime risk of posttraumatic stress disorder among abused and neglect children grown up. *Journal of Traumatic Stress*, 22, 566-574. doi: 10.1002/jts.20478
- \*Kugler, B. B., Bloom, M., Kaercher, L. B., Truax, T. V., & Storch, E. A. (2012). Somatic symptoms in traumatized children and adolescents. *Child Psychiatry & Human Development*, 43, 661-673. doi:10.1007/s10578-012-0289-y
- \*Kuo, J. R., Goldin, P. R., Werner, K., Heimberg, R. G., & Gross, J. J. (2011). Childhood trauma and current psychological functioning in adults with social anxiety disorder. *Journal of Anxiety Disorders*, 25, 467-473. doi:10.1016/j.janxdis.2010.11.011
- \*Lamoureux, B. E., Palmierir, P. A., Jackson, A. P., & Hobfoll, S. E. (2012). Child sexual abuse and adulthood interpersonal outcomes: Examining pathways for intervention. *Psychological Trauma: Theory, Research, Practice, and Policy*, 4, 605-6013. doi: 10.1037/a0026079
- \*Leck, P., Difede, J., Patt, I., Giosan, C., & Szkodny, L. (2006). Incidence of male childhood sexual abuse and psychological sequelae in disaster workers exposed to a terrorist attack. *International Journal of Emergency Mental Health*, 8, 267-274.
- \*Li, N., Ahmed, S., & Zabin, L. S. (2012). Association between childhood sexual abuse and adverse psychological outcomes among youth in Taipei. *Journal of Adolescent Health*, 50, S45-S51. doi: 10.1016/j.jadohealth.2011.12.003
- \*Liem, J. H., O'Toole, J. G., & James, J. B. (1996). Themes of power and betrayal in sexual abuse survivors' characterizations of interpersonal relationship. *Journal of Traumatic Stress*, 9, 745-761.
- \*López, F., Carpintero, E., Hernández, A., Martín, M. J., & Fuertes, A. (1995). Prevalencia y consecuencias del abuso sexual al menor en España [Prevalence and consequences of sexual abuse in children in Spain]. *Child Abuse & Neglect*, 19, 1039-1050. doi: 10.1016/0145-2134(95)00066-H

- \*Lumley, M. N., & Harkness, K. L. (2007). Specificity in the relations among childhood adversity, early maladaptive schemas, and symptom profiles in adolescent depression. *Cognitive Therapy and Research, 31*, 639-657. doi: 10.1007/s10608-006-9100-3
- \*Luterek, J. A., Harb, G. C., Heimberg, R. G., & Marx, B. P. (2004). Interpersonal rejection sensitivity in childhood sexual abuse survivors. Mediator of depressive symptoms and anger suppression. *Journal of Interpersonal Violence, 19*, 90-107. doi: 10.1177/0886260503259052
- \*Lynskey, M. T., & Fergusson, D. M. (1997). Factors protecting against the development of adjustment difficulties in young adults exposed to childhood sexual abuse. *Child Abuse & Neglect, 21*, 1177-1190. doi: 10.1016/S0145-2134(97)00093-8
- \*McMillan, H. L., Fleming, J. E., Streiner, D. L., Lin, E., Boyle, M. H., Jamieson, E.,...Beardslee, W. R. (2001). Childhood abuse and lifetime psychopathology in a community sample. *American Journal of Psychiatry, 158*, 1878-1883. doi: 10.1176/appi.ajp.158.11.1878
- Maniglio, R. (2009). The impact of child sexual abuse on health: A systematic review of reviews. *Clinical Psychology Review, 29*, 647-657. doi: 10.1016/j.cpr.2009.08.003
- Maniglio, R. (2010). Child sexual abuse in the etiology of depression: A systematic review of reviews. *Depression and Anxiety, 27*, 631-642. doi: 10.1002/da.20687
- \*Manion, I., Firestone, P., Cloutier, P., Ligezinska, M., McIntyre, J., & Ensom, R. (1998). Child extrafamilial sexual abuse: Predicting parent and child functioning. *Child Abuse & Neglect, 22*, 1285-1304. doi: 10.1016/S0145-2134(98)00103-3
- \*Mapp, S. C. (2006). The effects of sexual abuse as a child on the risk of mothers physically abusing their children: A path analysis using systems theory. *Child Abuse & Neglect, 30*, 1293-1310. doi: 10.1016/j.chiabu.2006.05.006
- McGraw, K. O., & Wong, S. P. (1992). A common language effect size statistic. *Psychological Bulletin, 111*, 361-365. doi:10.1037/0033-2909.111.2.361
- \*Mchichi Alami, K., & Kadri, N. (2004). Moroccan women with a history of child sexual abuse and its long-term repercussions: A population-based epidemiological study. *Archives of Women's Mental Health, 7*, 237-242. doi
- \*McLean, C. P., Morris, S. H., Conkin, P., Jayawichreme, N., & Foa, E. B. (2014). Trauma characteristics and posttraumatic stress disorder among adolescent survivors of childhood sexual abuse. *Journal of Family Violence, 29*, 559-566. doi: 10.1007/s10896-014-9613-6
- \*McLeer, S. V., Dixon, J. F., Henry, D., Ruggiero, K., Escovitz, K., Niedda, T., & Scholle, R. (1998). Psychopathology in non-clinically referred sexually abused children. *Journal of the American Academy of Child & Adolescent Psychiatry, 37*, 1326-1333. doi: 10.1097/00004583-199812000-00017
- Mennen, F. E. (1995). The relationship of race/ethnicity to symptoms in childhood sexual abuse. *Child Abuse & Neglect, 19*, 115-124. doi: 10.1016/0145-2134(94)00100-9
- \*Merril, L. L. (2001). Trauma symptomatology among female U.S. Navy recruits. *Military Medicine, 166*, 621-624.
- \*Messman-Moore, T. L., Long, P. J., & Siegfried, N. J. (2000). The revictimization of child sexual abuse survivors: An examination of the adjustment of college women with child sexual abuse, adult sexual assault, and adult physical abuse. *Child Maltreatment, 5*, 18-27. doi: 10.1177/1077559500005001003
- \*Meston, C. M., Rellini, A. H., & Heiman, J. R. (2006). Women's history of sexual abuse, their sexuality, and sexual self-schemas. *Journal of Consulting and Clinical Psychology, 74*, 229-236. doi: 10.1037/0022-006X.74.2.229

- \*Meyerson, L. A., Long, P. J., Miranda, R., Jr., & Marx, B. P. (2002). The influence of childhood sexual abuse, physical abuse, family environment, and gender on the psychological adjustment of adolescents. *Child Abuse & Neglect*, *26*, 387–405. doi: 10.1016/S0145-2134(02)00315-0
- \*Miller, D. K. (2006). The effects of childhood physical abuse or childhood sexual abuse in battered women's coping mechanisms: Obsessive-compulsive tendencies and severe depression. *Journal of Family Violence*, *21*, 185-195. doi: 10.1007/s10896-006-9019-1
- \*Molnar, B. E., Buka, S. L., & Kessler, R. C. (2001). Child sexual abuse and subsequent psychopathology: Results from the National Comorbidity Survey. *American Journal of Public Health*, *91*, 753–760. doi: 10.2105/AJPH.91.5.753
- \*Mullen, P. E., Martin, J. L., Anderson, J. C., Romans, S. E., & Herbison, G. P. (1996). The long-term impact of the physical, emotional, and sexual abuse of children: A community study. *Child Abuse & Neglect*, *20*, 7–21. doi: 10.1016/0145-2134(95)00112-3
- \*Musliner, K. L., & Singer, J. B. (2014). Emotional support and adult depression in survivors of childhood sexual abuse. *Child Abuse & Neglect*, *38*, 1331-1340. doi: 10.1016/j.chiabu.2014.01.016
- \*Nelson, E. C., Heath, A. C., Madden, P. A. F., Cooper, M. L., Dinwiddie, S. H., Bucholz, K. K.,...Martin, N. G. (2002). Association between self-reported childhood sexual abuse and adverse psychosocial outcomes: Results from a twin study. *Archives of General Psychiatry*, *59*, 139-145. doi: 10.1001/archpsyc.59.2.139
- \*Newcomb, M. D., Munoz, D. T., & Vargas Carmona, J. (2009). Child sexual abuse consequences in community samples of Latino and European American adolescents. *Child Abuse & Neglect*, *33*, 533-544. doi: 10.1016/j.chiabu.2008.09.014
- Nolen-Hoeksema, S. (1990). *Sex differences in depression*. Stanford, CA: Stanford University Press.
- Nolen-Hoeksema, S. (2002). Gender differences in depression. In I. H. Gotlib & C. L. Hammen (Eds.), *Handbook of depression* (pp. 492-509). New York: Guilford Press.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York, NY: McGraw-Hill.
- \*Offen, L., Waller, G., & Thomas, G. (2003). Is reported childhood sexual abuse associated with the psychological characteristics of patients who experience auditory hallucinations? *Child Abuse & Neglect*, *27*, 919-927. doi: 10.1016/S0145-2134(03)00139-X
- Paolucci, E. O., Genuis, M. L., & Violato, C. (2001). A meta-analysis of the published research on the effects of child sexual abuse. *Journal of Psychology*, *135*, 17-36. doi: 10.1080/00223980109603677
- \*Peleikis, D. E., Mykletun, A., & Dahl, A. A. (2004). The relative influence of childhood sexual abuse and other family background risk factors on adult adversities in female outpatients treated for anxiety disorders and depression. *Child Abuse & Neglect*, *28*, 61-76. doi: 10.1016/j.chiabu.2003.08.003
- \*Peleikis, D. E., Mykletun, A., & Dahl, A. A. (2005). Current mental health in women with childhood sexual abuse who had outpatient psychotherapy. *European Psychiatry*, *20*, 260-267. doi: 10.1016/j.eurpsy.2005.01.004
- Pereda, N., Guilera, G., Forns, M., & Gómez-Benito, J. (2009). The prevalence of child sexual abuse in community and student samples: A meta-analysis. *Clinical Psychology Review*, *29*, 328-338. doi: 10.1016/j.cpr.2009.02.007

- \*Pérez-Fuentes, G., Olfson, M., Villegas, L., Morcillo, C., Wang, S., & Blanco, C. (2013). Prevalence and correlates of child sexual abuse: A national study. *Comprehensive Psychiatry*, *54*, 16-27. doi: 10.1016/j.comppsy.2012.05.010
- \*Portegijs, P. J. M., Jeuken, F. M. H., van der Horst, F. G., Kraan, H. F., & Knottnerus, J. A. (1996). A troubled youth: Relations with somatization, depression and anxiety in adulthood. *Family Practice*, *13*, 1-11. doi: 10.1093/fampra/13.1.1
- \*Rich, C. L., Gidycz, C. A., Warkentin, J. B., Loh, C., & Weiland, P. (2005). Child and adolescent abuse and subsequent victimization: A prospective study. *Child Abuse & Neglect*, *29*, 1373–1394. doi: 10.1016/j.chiabu.2005.07.003
- Rind, B., Bauserman, R., & Tromovich, P. (1998). A meta-analytic examination of assumed properties of child sexual abuse using college samples. *Psychological Bulletin*, *124*, 22-53. doi: 10.1037/0033-2909.124.1.22
- Rosenthal, R. (1994). Parametric measures of effect size. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 231–244). New York, NY: Russell Sage Foundation.
- \*Schaaf, K. K., & McCanne, T. R. (1998). Relationship of childhood sexual, physical, and combined sexual and physical abuse to adult victimization and posttraumatic stress disorder. *Child Abuse & Neglect*, *22*, 1119-1133. doi: 10.1016/S0145-2134(98)00090-8
- Schoedl, A. F., Costa, M. C. P., Mari, J. J., Mello, M. F., Tyrka, A. R., Carpenter, L. L., & Price, L. H. (2010). The clinical correlates of reported childhood sexual abuse: An association between age at trauma onset and severity of depression and PTSD in adults. *Journal of Child Sexual Abuse*, *19*, 156-170. doi: 10.1080/10538711003615038
- \*Silverman, A. B., Reinherz, H. A., & Giaconia, R. M. (1996). The long-term sequelae of child and adolescent abuse: A longitudinal community study. *Child Abuse & Neglect*, *20*, 709-723. doi: 10.1016/0145-2134(96)00059-2
- \*Spertus, I. L., Yehuda, R., Wong, C. M., Halligan, S., Seremetis, S. V. (2003). Childhood emotional abuse and neglect as predictors of psychological and physical symptoms in women presenting to a primary care practice. *Child Abuse & Neglect*, *27*, 1247-1258. doi: 10.1016/j.chiabu.2003.05.001
- \*Steel, J., Sanna, L., Hammond, B., Whipple, J., & Cross, H. (2004). Psychological sequelae of childhood sexual abuse: Abuse-related characteristics, coping strategies, and attributional style. *Child Abuse & Neglect*, *28*, 785-801. doi: 10.1016/j.chiabu.2003.12.004
- Stoltenborgh, M., van Ijzendoorn, M. H., Euser, E. M., & Bakermans-Kranenburg, M. J. (2011). A global perspective on child sexual abuse: Meta-analysis of prevalence around the world. *Child Maltreatment*, *16*, 79-101. doi: 10.1177/1077559511403920
- \*Subica, A. M. (2013). Psychiatric and physical sequelae of childhood physical and sexual abuse and forced sexual trauma among individuals with serious mental illness. *Journal of Traumatic Stress*, *26*, 588-596. doi: 10.1002/jts.21845
- \*Sun, Y.P., Zhang, B., Dong, Z.J., Yi, M. J., Sun, D. F., & Shi, S. S. (2008). Psychiatric state of college students with a history of childhood sexual abuse. *World Journal of Pediatrics*, *4*, 285-290. doi 10.1007/s12519-008-0052-4
- \*Swanston, H. Y., Plunkett, A. M., O’Toole, B. I., Shrimpton, S., Parkinson, P. N., & Oates, R. K. (2003). Nine years after child sexual abuse. *Child Abuse & Neglect*, *27*, 967–984. doi:10.1016/s0145-2134(03)00143-1
- \*Thomas, R. A., DiLillo, D., Walsh, K., & Polusny, M. A. (2011). Pathways from child sexual abuse to adult depression: The role of parental socialization of emotions and alexithymia. *Psychology of Violence*, *1*, 121-135. doi: 10.1037/a0022469

- \*Thompson, K. M., Crosby, R. D., Wonderlich, S. A., Mitchell, J. E., Redlin, J., Demuth, G., Smyth, J., & Haseltine, B. (2003). Psychopathology and sexual trauma in childhood and adulthood. *Journal of Traumatic Stress, 16*, 35-38. doi: 10.1023/A:1022007327077
- Tolin, D. F., & Foa, E. B. (2006). Sex differences in trauma and posttraumatic stress disorder: A quantitative review of 25 years of research. *Psychological Bulletin, 132*, 959-992. doi: 10.1037/0033-2909.132.6.959
- \*Trowell, J., Ugarte, B., Kolvin, I., Berelowitz, M., Sadowski, H., & Le Couteur, A. (1999). Behavioural psychopathology of child sexual abuse in schoolgirls referred to a tertiary centre: A north London study. *European Child & Adolescent Psychiatry, 8*, 107-116. doi: 10.1007/s007870050091
- United Nations. (1988). *Committee on crime prevention and control. Report on the tenth session*. Vienna: United Nations Publications.
- \*van Vugt, E., Lanctôt, N., Paquette, G., Collin-Vézina, D., & Lemieux, A. (2014). Girls in residential care: From child maltreatment to trauma-related symptoms in emerging adulthood. *Child Abuse & Neglect, 38*, 114-122. doi: 10.1016/j.chiabu.2013.10.015
- Vilariño, M., Arce, R., & Fariña, F. (2013). Forensic-clinical interview: Reliability and validity for the evaluation of psychological injury. *The European Journal of Psychology Applied to Legal Context, 5*, 1-21.
- \*Villarroel, A. M., Penelo, E., Portell, M., & Raich, R. M. (2012). Childhood sexual and physical abuse in Spanish female undergraduates: Does it affect eating disturbances? *European Eating Disorders Review, 20*, e32-e41. doi: 10.1002/erv.1086
- \*Widom, C. S., DuMont, K., & Czaja, S. J. (2007). A prospective investigation of major depressive disorder and comorbidity in abused and neglected children grown up. *Archives of General Psychiatry, 64*, 49-56. doi: 10.1001/archpsyc.64.1.49
- World Health Organization. (2000). *Women's mental health: An evidence based review*. Geneva, Switzerland: Author.
- World Health Organization. (1999, March 29-31). *Report of the consultation on child abuse prevention*. Geneva, Switzerland: Author. Retrieved from <http://www.who.int/iris/handle/10665/65900#sthash.YgBC2YVe.dpuf>
- World Health Organization. (2014, December). *Child Maltreatment* [Fact sheet N° 150]. Retrieved from <http://www.who.int/mediacentre/factsheets/fs150/es/>
- \*Young, M. S., Harford, K. L., Kinder, B., & Savell, J. K. (2007). The relationship between childhood sexual abuse and adult mental health among undergraduates: Victim gender doesn't matter. *Journal of Interpersonal Violence, 22*, 1315-1331. doi: 10.1177/0886260507304552

Appendix 1. Summary Table of Primary Studies Characteristics

	$N_{GE}$	$N_{GC}$	$r$	$r_{xx}$	$r_{yy}$	CSA Questionnaire	Depression/Anxiety Measure	Type of measure
Balsam, Lehavot, Beadnell, & Circo (2010)	669	669	.13	.94	.86	CTQ-SF < 17 years	PHQ GAD-7 CESD-10	Anxiety symptomatology Depressive symptomatology
Bonomi, Cannon, Anderson, Rivara, & Thompson (2008) <sup>a</sup>	693	2399	.05	-	.76	Behavioral Risk Factor Surveillance System < 18 years	CES-D	Depressive symptomatology
Briere and Elliot (2003) <sup>a</sup>	152	309	.17	-	.86	TES < 18 years	TSI	Anxiety symptomatology Depressive symptomatology
Briere and Elliot (2003) <sup>b</sup>	66	398	.26	-	.86	TES < 18 years	TSI	Anxiety symptomatology Depressive symptomatology
Brown, Cohen, Johnson, & Smailes (1999)	22	558	.72	-	-	Forensic Sample Forensic Sample	DSM-III-R DSM-III-R	Dysthymia Major Depressive Disorder
Cantón-Cortés, Cortés, & Cantón (2012)	182	182	.30	-	.86	Childhood sexual abuse questionnaire < 13 years	BDI STAI	Depressive symptomatology Anxiety symptomatology
Cantón-Cortés & Justicia (2008)	83	83	.22	-	.86	Questionnaire < 13 years	BDI	Depressive symptomatology
Carey, Walker, Rossouw, Seedat, & Stein (2008)	50	44	.03	-	-	CTQ < 17 years	DSM-IV DSM-IV DSM-IV DSM-IV	Major Depressive Disorder Dysthymia Panic Disorder Social Phobia
Cheasty, Clare, & Collins (1998) <sup>a</sup>	73	164	.15	-	.86	30-item general Health questionnaire < 16 years	BDI	Depressive symptomatology
Chen, Dunne, & Han (2004) <sup>a</sup>	89	944	.16 <sup>(1)</sup>	-	.89	Questionnaire < 16 years	CES-D	Depressive symptomatology
	102	944	.24 <sup>(2)</sup>	-	.89	< 16 years		Depressive symptomatology

	$N_{GE}$	$N_{GC}$	$r$	$r_{xx}$	$r_{yy}$	CSA Questionnaire	Depression/Anxiety Measure	Type of measure
Chen, Dunne, & Han (2004) <sup>b</sup>	62	990	.14 <sup>(1)</sup>	-	.89	Questionnaire	CES-D	Depressive symptomatology
	55	990	.16 <sup>(2)</sup>	-	.89	<16 years		Depressive symptomatology
Chen, Dunne, & Han (2006) <sup>a</sup>	28	269	.08 <sup>(1)</sup>	-	.91	Questionnaire	CES-D	Depressive symptomatology
	49	269	.24 <sup>(2)</sup>	-	.91	< 16 years		Depressive symptomatology
Chen et al. (2014) <sup>a, (1)</sup>	233	10882	.30	-	-	Stressful life events	CIDI ( DSM-IV)	Major Depressive Disorder
Chen et al. (2014) <sup>a, (2)</sup>	281	10834	.58	-	-	Stressful life events	CIDI ( DSM-IV)	Major Depressive Disorder
Chen et al. (2014) <sup>a, (3)</sup>	171	10944	.74	-	-	Stressful life events	CIDI ( DSM-IV)	Major Depressive Disorder
Comijs et al. (2013)	85	175	.23	-	-	Structured inventory	IDS	Depressive symptomatology
	85	175	.14	-	-	< 16 years	BAI	Anxiety symptomatology
Cortés-Arboleda, Cantón-Cortés, & Cantón-Duarte (2011) <sup>a</sup>	209	209	.20	-	.92	Questionnaire	STAI	Anxiety symptomatology
						< 18 years		
Cortés-Arboleda, Cantón-Duarte, & Cantón-Cortés (2011) <sup>a</sup>	240	240	.22	-	.92	Questionnaire	STAI	Anxiety symptomatology
	240	240	.29	-	.83	< 16 years	BDI	Depressive symptomatology
Cortés-Arboleda, Cantón-Duarte, & Cantón-Cortés (2011) <sup>b</sup>	29	29	.29	-	.92	Questionnaire	STAI	Anxiety symptomatology
	29	29	.18	-	.83	< 16 years	BDI	Depressive symptomatology
Cutajar et al. (2010) <sup>a</sup>	2153	2055	.34	-	-	Forensic Sample	ICD	Depressive Disorder
	2153	2055	.36	-	-	< 16 years	ICD	Anxiety Disorder
Cutajar et al. (2010) <sup>b</sup>	535	622	-.02	-	-	Forensic Sample	ICD	Depressive Disorder
	535	622	.40	-	-	< 16 years	ICD	Anxiety Disorder
Doerfler, Toscano Jr., & Connor (2009)	39	73	.22	-	.91	Forensic Sample	DSMD	Depressive symptomatology
	39	73	.27	-	.86		DSMD	Anxiety symptomatology
Dube et al. (2005) <sup>a</sup>	1173	3520	.13	.69	-	ACE	Screening Instrument for depressive disorders	Major Depressive Disorder
Dube et al. (2005) <sup>b</sup>	601	3414	.07	.69	-	ACE	Screening Instrument for depressive disorders	Major Depressive Disorder
						< 18 years		
						< 18 years		

	$N_{GE}$	$N_{GC}$	$r$	$r_{xx}$	$r_{yy}$	CSA Questionnaire	Depression/Anxiety Measure	Type of measure
Feeney, Kamiya, Robertson, & Kenny (2013)	451	6256	.08	-	.85	2 questions	CE5-D	Depressive symptomatology
	451	6256	.08	-	.80	< 18 years	HADS-A	Anxiety symptomatology
Feerick & Snow (2005) <sup>a</sup>	98	215	.13	-	.84	CSAI < 18 years	H5CL	Anxiety symptomatology
Fergusson, Boden, & Horwood (2008) <sup>(1)</sup>	28	881	.09	-	-	Interview	CIDI	Anxiety Disorder
	28	881	.12	-	-	<16 years	DSM-IV	Major Depressive Disorder
Fergusson, Boden, & Horwood (2008) <sup>(2)</sup>	52	881	.17	-	-	Interview	CIDI	Anxiety Disorder
	52	881	.23	-	-	< 16 years	DSM-IV	Major Depressive Disorder
Fergusson, Boden, & Horwood (2008) <sup>(3)</sup>	64	881	.19	-	-	Interview	CIDI	Anxiety Disorder
	64	881	.27	-	-	<16 years	DSM-IV	Major Depressive Disorder
Ferguson & Dacey (1997) <sup>a</sup>	19	55	.41	-	.85	CEQ	BDI	Depressive symptomatology
	19	55	.37	-	.90	-	STAI	Anxiety symptomatology
Fergusson, McLeod, & Horwood (2013)	28	809	.08 <sup>(1)</sup>	-	-	Structured	CIDI	Major Depressive Disorder
	51	809	.15 <sup>(2)</sup>	-	-	Interview	-	Major Depressive Disorder
	62	809	.20 <sup>(3)</sup>	-	-	< 16 years	-	Major Depressive Disorder
	28	809	.05 <sup>(1)</sup>	-	-	Structured	CIDI	Anxiety Disorder
	51	809	.09 <sup>(2)</sup>	-	-	Interview	-	Anxiety Disorder
	62	809	.22 <sup>(3)</sup>	-	-	< 16 years	-	Anxiety Disorder
Fondacaro, Holt, & Powell (1999) <sup>b</sup>	86	125	.18	-	-	Questionnaire	DIS (DSM-III-R)	Major Depressive Disorder
	86	125	.06	-	-	< 16 years	-	Dysthymia
	86	125	.23	-	-	-	-	Panic Disorder
	86	125	.25	-	-	-	-	Generalized Anxiety Disorder
Frias, Brassard, & Shaver (2014) <sup>a</sup>	116	691	.10	-	.90	1 question	ECR	Anxiety symptomatology
Godbout, Briere, Sabourin, & Luissier (2013)	59	284	.23	-	.88	SCEQ	ECR	Anxiety symptomatology
						<18 years		
Gudjonsson, Sigurdsson, & Tryggvadóttir (2011)	37	73	.12	-	.84	Parental Neglect	DASS	Anxiety symptomatology
	37	73	.06	-	.91	and Sexual Abuse	DASS	Depressive symptomatology
Haj-Yahia & Tamish (2001)	652		.55	.89	.88	Sexual Abuse	BSI	Anxiety symptomatology
	652		.60	.89	.88	Finkelhor's	BSI	Depressive symptomatology
	652		.59	.89	-	Scale	-	Anxiety symptomatology
Henderson, Hargreaves, Gregory, & Williams (2002) <sup>a</sup>	22	57	.27	-	-	Interview	POMS-SF	Depressive symptomatology
	22	57	.31	-	-	<14 years	POMS-SF	Anxiety symptomatology
Hobfoll et al. (2002)	67		.05	.93	.90	CTQ	POMS	Depressive symptomatology
						< 17 years		



Jonas et al. (2011)	964	6389	.21	-	.75	Interview	CIS-R	Major Depressive Disorder
		.19	-	.75	<16 years		CIS-R	Generalized Anxiety Disorder
			.18	-	.75		CIS-R	Panic Disorder
			.28	-	.75		CIS-R	Phobic Disorder
Kendler et al. (2000)	427	983	.24	-	-	Interview	SCI (DSM-III-R)	Generalized Anxiety Disorder
	427	983	.24	-	-	< 16 years	SCI	Panic disorder
	427	983	.25	-	-		SCI	Major Depressive Disorder
Kent & Waller (1998) <sup>a</sup>	236		.30	.61	.70	CATS	HADS-A	Anxiety symptomatology
	236		.28	.61	.60		HADS-D	Depressive symptomatology
Kugler, Bloom, Kaercher, Truax, & Storch	54	107	.57	-	.88	Forensic sample	TSCC	Anxiety symptomatology
	54	107	.62	-	.81	8-17 years	CDI and TSCC	Depressive symptomatology
Kuo, Goldin, Werner, Heimberg, & Gross (2011)	20	82	.10	.86	.93	CTQ-SF	SIAS	Anxiety symptomatology
	20	82	.07	.86	.90	< 16 years	BDI-II	Depressive symptomatology
Lamoureux, Palmieri, Jackson, & Hobfoll (2012) <sup>a</sup>	271	422	.26	.87	.89	CTQ	CES-D	Depressive symptomatology
						<16 years		
Leck, Difede, Patt, Giosan, & Szkodny (2006) <sup>b</sup>	92	2030	.18	.83	.93	TEI	BDI-II	Depressive symptomatology
Li, Ahmed, & Zabin (2012)	214	3870	.27	-	-	Research Study of	1 question	Anxiety symptomatology
	214	3870	.28	-	-	Adolescent Health	1 question	Depressive symptomatology
						<14 years		
Liem, O'Toole, & James (1996) <sup>a</sup>	43	43	.24	-	.81	SEQ	BSI	Anxiety symptomatology
	.25	.85	<14 years	BSI				43
	43	43	.29	-	.83		BDI-SF	43
Linskey & Fergusson (1997)	24	918	.11 <sup>(1)</sup>	-	-	Reports of	DSM-IV	Depressive symptomatology
	47	918	.15 <sup>(2)</sup>	-	-	Childhood Sexual	DSM-IV	Anxiety Disorder
	36	918	.15 <sup>(3)</sup>	-	-	abuse	DSM-IV	Anxiety Disorder
	24	918	.12 <sup>(1)</sup>	-	-	<16 years	DSM-IV	Depressive Disorder
	47	918	.17 <sup>(2)</sup>	-	-		DSM-IV	Depressive Disorder
	36	918	.21 <sup>(3)</sup>	-	-		DSM-IV	Depressive Disorder
López, Carpintero, Hernández, Martín, & Fuertes (1995)	337	1484	.12	-	-	Interview	SRQ	Anxiety symptomatology
	337	1484	.08	-	-	< 16 years	SRQ	Depressive symptomatology
Lumley & Harkness (2007)	11		.13	.93	.91	CECA	MASQ	Anxiety symptomatology
	11		.13	.93	.88		MASQ	Depressive symptomatology
Luterek, Harb, Heimberg, & Marx (2004) <sup>a</sup>	34		.12	-	.81	LEQ < 14 years	BDI	Depressive symptomatology
	<b>N<sub>GE</sub></b>	<b>N<sub>GC</sub></b>	<b>r</b>	<b>r<sub>xx</sub></b>	<b>r<sub>yy</sub></b>	<b>CSA Questionnaire</b>	<b>Depression/Anxiety Measure</b>	<b>Type of measure</b>
MacMillan et al. (2001) <sup>a</sup>	508	3170	.08	-	-	Child Maltreatment	CIDI	Anxiety Disorder
	508	3170	.07	-	-	History Self-Report	CIDI	Depressive Disorder
MacMillan et al. (2001) <sup>b</sup>	150	3188	.02	-	-	Child Maltreatment	CIDI	Anxiety Disorder

	150	3188	.02	-	-	History Self-Report	CIDI	Depressive Disorder
Manion et al. (1998) <sup>a</sup>	29	45	.47	-	-	NAEF <14years	Depression Self-Rating Scale for children	Depressive symptomatology
Manion et al. (1988) <sup>b</sup>	22	29	.39	-	-	NAEF <14years	Depression Self-Rating Scale for children	Depressive symptomatology
Mapp (2006) <sup>a</sup>	107	158	.13	-	.87	Forensic Sample <18 years	CES-D	Depressive symptomatology
Mchichi Alami & Kadri (2004) <sup>a</sup>	62	620	.09	-	-	Questionnaire	Hamilton depression rating scale	Depressive symptomatology
	19	620	.10 <sup>(1)</sup>	-	-			Depressive symptomatology
	21	620	.03 <sup>(2)</sup>	-	-			Depressive symptomatology
	22	620	.03 <sup>(3)</sup>	-	-			Depressive symptomatology
	63	617	.02	-	-		Hamilton anxiety rating scale	Anxiety symptomatology
	21	617	.05 <sup>(1)</sup>	-	-			Anxiety symptomatology
	20	617	.01 <sup>(2)</sup>	-	-			Anxiety symptomatology
	22	617	.01 <sup>(3)</sup>	-	-			Anxiety symptomatology
McLean, Morris, Conklin, Jayawickreme, & Foa (2014) <sup>a</sup>	71		.59	-	.87	Forensic Sample 13-18 years	BDI	Depressive symptomatology
McLeer et al. (1998)	80	73	.33	-	.86	Interview	CDI	Depressive symptomatology
	80	73	.27	-	.89	6-16	STAIC	Anxiety symptomatology
Merril (2001) <sup>a</sup>	248	523	.15	-	.84	SEQ	TSI	Anxiety symptomatology
	248	523	.17	-	.84	<14 years	TSI	Depressive symptomatology
Messman-Moore, Long, & Siegfried (2000) <sup>a</sup>	56	282	.24	.89	.85	LEQ	SCL-90-R	Anxiety symptomatology
	56	282	.17	.89	.90	< 17 years	SCL-90-R	Depressive symptomatology
Meyerson, Long, Miranda, & Marx (2002)	39	91	.23	.75	.93	SEQ <12 years	BDI-II	Depressive symptomatology
Meston, Rellini, & Heiman (2006) <sup>a</sup>	48	71	.34	-	.92	Questionnaire	BAI	Anxiety symptomatology
	48	71	.40	-	.86	< 16 years	BDI	Depressive symptomatology
Miller (2006) <sup>a</sup>	25	50	.22	-	.86	Interview	BDI	Depressive symptomatology
	<b>N<sub>GE</sub></b>	<b>N<sub>GC</sub></b>	<b>r</b>	<b>r<sub>xx</sub></b>	<b>r<sub>yy</sub></b>	<b>CSA Questionnaire</b>	<b>Depression/Anxiety Measure</b>	<b>Type of measure</b>
Molnar, Buka, & Kessler (2001) <sup>a</sup>	394	2527	.13	-	-	<18 years	DSM-III-R	Generalized Anxiety Disorder
	394	2527	.13	-	-		DSM-III-R	Panic Disorder
	394	2527	.13	-	-		DSM-III-R	Phobic Disorder
	394	2527	.23	-	-		DSM-III-R	Major Depressive Disorder
	394	2527	.25	-	-		DSM-III-R	Dysthymia

Molnar, Buka, & Kessler (2001) <sup>b</sup>	74	2871	.04	-	-	<18 years	DSM-III-R	Generalized Anxiety Disorder
	74	2871	.09	-	-		DSM-III-R	Panic Disorder
	74	2871	.18	-	-		DSM-III-R	Phobic Disorder
	74	2871	.23	-	-		DSM-III-R	Major Depressive Disorder
	74	2871	.16	-	-		DSM-III-R	Dysthymia
Mullen, Martin, Anderson, Romans, & Herbison (1996) <sup>a</sup>	53	390	.21	-	-	Questionnaire <16 years	PSE-SF	Depressive symptomatology
Musliner & Singer (2014) <sup>a</sup>	436	221	.24	-	.85	Questionnaire <16 years	CE5-D-10	Depressive symptomatology
Nelson et al. (2002) <sup>a</sup>	387	1931	.12	-	-	Interview	DSM-IV	Social Phobia
	387	1931	.17	-	-	<18 years	DSM-IV	Major Depressive Disorder
Nelson et al. (2002) <sup>b</sup>	90	1574	.04	-	-	Interview	DSM-IV	Social Phobia
	90	1574	.08	-	-	<18 years	DSM-IV	Major Depressive Disorder
Newcomb, Munoz, & Carmona (2009) <sup>a</sup>	66	79	.26	-	.77	CMIS-SF	TSI	Anxiety symptomatology
	66	79	.20	-	.86	<17 years	TSI	Depressive symptomatology
Newcomb, Munoz, & Carmona (2009) <sup>b</sup>	19	59	.32	-	.77	CMIS-SF	TSI	Anxiety symptomatology
	19	59	.35	-	.86	<17 years	TSI	Depressive symptomatology
Offen, Waller, & Thomas (2003)	10	16	.42	-	-	1 Question	BDI	Depressive Symptomatology
Peteikis, Mykletun, & Dahl (2004) <sup>a</sup>	56	56	.52	-	-	Interview	SCID-II	Major Depressive Disorder
	56	56	.27	-	-	<16 years		Dysthymia

	$N_{GE}$	$N_{GC}$	$r$	$r_{xx}$	$r_{yy}$	CSA Questionnaire	Depression/Anxiety Measure	Type of measure
Peleikis, Mykletun, & Dahl (2005) <sup>a</sup>	56	56	0	-	-	Detailed Structured	SCID	Panic Disorder
	56	56	.10	-	-	Interview	SCID	Agoraphobia
	56	56	.07	-	-	<16 years	SCID	Social Phobia
	56	56	.24	-	-		SCID	Generalized Anxiety Disorder
	56	56	.12	-	.85		SCL-90-R	Anxiety symptomatology
	56	56	.14	-	.82		SCL-90-R	Anxiety symptomatology
	56	56	.19	-	-		SCID	Major Depressive Disorder
	56	56	.23	-	-		SCID	Dysthymia
	56	56	.18	-	.90		SCL-90-R	Depressive symptomatology
	Pérez-Fuentes et al. (2013)	3786	30431	.45	-	-	ACE	DSM-IV
3786		30431	.39	-	-	<18 years	DSM-IV	Social Phobia
3786		30431	.34	-	-		DSM-IV	Specific Phobia
3786		30431	.44	-	-		DSM-IV	Generalized Anxiety Disorder
3786		30431	.35	-	-		DSM-IV	Major Depressive Disorder
			.40	-	-		DSM-IV	Dysthymia
Portegijs, Jeuken, van der Horst, Kraan, & Knottnerus (1996) <sup>a</sup>	11		.07	-	-	Youth Experiences	DIS	Anxiety Disorder
	11		.33	-	-	Questionnaire	DIS	Depressive Disorder
Rich, Gidycz, Warkentin, Loh, & Weiland (2005) <sup>c</sup>	42		.09	-	.84	Child sexual victimization Quest.	BDI-II	Depressive symptomatology
				-		<14 years		
Rich, Gidycz, Warkentin, Loh, & Weiland (2005) <sup>d</sup>	189		.28	-	.74	SES	BDI-II	Depressive symptomatology
				-		<18 years		
Schaaf & McCanne (1998) <sup>a</sup>	27	211	.06	-	-	CSEQ	TSI	Anxiety symptomatology
	27	211	.11	-	-	< 15 years	TSI	Depressive symptomatology
Silverman, Reinherz, & Giaconia (1996) <sup>a</sup>	23	164	.08	-	-	Interview	DIS (DSM-III-R)	Specific Phobia
	23	164	.03	-	-	<18 years	DIS	Social Phobia
	23	164	.23	-	-		DIS	Major Depressive Disorder
Spertus, Yehuda, Wong, Halligan, & Seremetis (2003) <sup>a</sup>	41	162	.20	.82	.85	CTQ	SCL-90-R	Anxiety symptomatology
	41	162	.18	.82	.90	< 17 years	SCL-90-R	Depressive symptomatology

	$N_{GE}$	$N_{GC}$	$r$	$r_{xx}$	$r_{yy}$	CSA Questionnaire	Depression/Anxiety Measure	Type of measure
Steel, Sanna, Hammond, Whipple, & Cross (2004)	85	172	.21	-	.85	Sexual History Questionnaire <18 years	SCL-90-R SCL-90-R	Anxiety symptomatology Depressive symptomatology
Subica (2013)	50	122	.25	-	.86	TAA-R <18 years	PHQ-8	Major Depressive Disorder
Sun et al. (2008)	244	781	.29	-	.85	Questionnaire <18 years	SCL-90-R SCL-90-R	Anxiety symptomatology Depressive symptomatology
Swanston et al. (2003)	104	63	.41	-	.82	Forensic Sample 5-15 years	RCMAS CDI	Anxiety symptomatology Depressive symptomatology
Thomas, DiLillo, Walsh, & Polusny (2011) <sup>a</sup>	52		.36	.85	.93	WSHQ <14 years	BDI-II	Depressive symptomatology
Thompson et al. (2003) <sup>a</sup>	26	25	.45	-	-	Interview <18 years	SCID-I SCID-I	Anxiety Disorder Depressive Disorder
Trowell et al. (1999) <sup>a</sup>	21	21	.48	-	-	Forensic Sample 8-14 years	Kiddie-SADS Kiddie-SADS	Major Depressive Disorder Generalized Anxiety Disorder Social Phobia Specific Phobia
van Vugt, Lanctôt, Paquette, Collin-Vézina, & Lemieux (2013) <sup>a</sup>	89	89	.46	.87	.88	CTQ	TSI-2 TSI-2	Anxiety symptomatology Depressive symptomatology
Villarroel, Penelo, Portell, & Raich (2012) <sup>a</sup>	81	597	.04	-	.93	TLEQ <18 years	STAI BDI	Anxiety symptomatology Depressive symptomatology
Widom, DuMont, & Czaja (2007)	96	520	.03	-	-	Forensic Sample <12 years	DIS	Major Depressive Disorder
Young, Harford, Kinder, & Savell (2007) <sup>a</sup>	116	163	.22	.79	.83	ESE	BSI BSI	Anxiety symptomatology Depressive symptomatology
Young, Harford, Kinder, & Savell (2007) <sup>b</sup>	39	88	.05	.79	.83	ESE <16 years	BSI BSI	Anxiety symptomatology Depressive symptomatology

Note.  $N_{GE}$  = experimental group sample size;  $N_{GC}$  = control group sample size;  $r$  = sexual abuse victimization and depression/anxiety correlation;  $r_{xx}$  = reliability of sexual abuse measure instruments;  $r_{yy}$  = reliability of Anxiety and Depression measure instruments; <sup>a</sup> = female participants; <sup>b</sup> = male participants; <sup>c</sup> = childhood sexual abuse (CSA); <sup>d</sup> = adolescence sexual abuse (AA).

(1) = Non-contact CSA; (2) = contact CSA; (3) = Intercourse  
 CTQ-SF = Childhood Trauma Questionnaire Short Form; TES = Traumatic Events Survey; CESD-10 = 10-item Center for Epidemiologic Studies Depression; PHQ-GAD-7 = 7-item Patient Health Questionnaire Generalized Anxiety Disorder Scale; CES-D = 20-item Center for Epidemiological Studies-Depression scale; BDI = Beck Depression Inventory; CIDI = Composite International Diagnostic Interview; IDS = Inventory of Depression Symptoms; STAI = State Trait Anxiety Inventory; ICD = International Classification of Disease; STAI = State Trait Anxiety Inventory for Children; DSMD = Devereux Scales of Mental Disorders; ACE = Adverse Childhood Experiences; HADS-A = Hospital Anxiety and Depression Scale - Anxiety Scale; CSAI = Childhood Sexual Abuse Interview; HSCL = Hopkins Symptom Checklist; CEQ = Childhood Experiences Questionnaire; ECR =

Experiences in Close Relationships Questionnaire; DASS = Depression Anxiety Stress Scales; BSI = Brief Symptom Inventory; POMS-SF = Profile of Mood States-Short Form; CTQ = Childhood Trauma Questionnaire; CIS-R = Clinical Interview Schedule Revised; CATS = Child Abuse and Trauma Scale; TSCC = Trauma Symptom Checklist for Children; CDI = Children Depression Inventory; TEI = Traumatic Events Interview; BDI-II = Beck Depression Inventory, second edition; TEQ = The Traumatic Events Questionnaire; BDI-SF = Beck Depression Inventory- Short Form; BSI = Brief Symptom Inventory; SEQ = The Significant Events Questionnaire; SRQ = Self Reporting Questionnaire; CECA = Childhood Experience of Care and Abuse Interview; MASQ = Mood and Anxiety Symptom Questionnaire; AA = Anxiety Arousal; TSI = Trauma Symptom Inventory; LEQ = Life Experiences Questionnaire; BAI = Beck Anxiety Inventory; PSE-SF = Present State Examination- Short Form; CMIS-SF = Childhood Maltreatment Interview Schedule-Short Form; SCID-II = Structured Clinical Interview for DSM-IV Axis I; CSEQ = Childhood Sexual Experiences Questionnaire; TAA-R = Trauma Assessment for Adults Brief Revised Version; PHQ-8 = Patient Health Questionnaire - 8; RCWAS = Revised Children's Manifest Anxiety Scale; WSHQ = CSA subscale from the Wyatt Sex History Questionnaire; Kiddie-SADS = Semi-Structured Interview Kiddie-Sads, DSM-IV; SICE = Structured Interview on CSA Experiences; TSC-33 = Trauma Symptom Checklist; ESE = Early Sexual Experiences; NIMH = National Institute of Mental Health Diagnostic Interview Schedule, Version III Revised; SES = Sexual Experiences Survey; SCL = Structured Clinical Interview; SIAS = Social Interaction Anxiety Scale; NAEF = Natural of the Abusive Experience Form; DIS = Diagnostic Interview Schedule; EPDS = Edinburgh Post-Natal Depression Scale; CCEI = Crown-Crisp Experiential Index; TLEQ = Traumatic Life Events Questionnaire; SCEQ = Childhood Sexual Experiences Questionnaire; RADS = Reynold's Adolescent Depression Scale.



#### 8.4. ANÁLISIS DE CONTENIDO EN DECLARACIONES DE AGRESORES. UNA REVISIÓN META-ANALÍTICA

Bárbara G. Amado, Manuel Vilariño\*, Mercedes Novo

Departamento de Psicología Organizacional, Jurídica-Forense y Metodología de las Ciencias del Comportamiento, Universidade de Santiago de Compostela (España)

\*Escuela Universitaria de Enfermería de Pontevedra, Universidade de Vigo (España)

##### Resumen

El análisis de contenido de la declaración, es una de las metodologías utilizadas en el contexto judicial para la evaluación de la credibilidad del testimonio. Concretamente, el instrumento de evaluación más ampliamente utilizado es el Criteria-Based Content Analysis (CBCA), que se ha aplicado originariamente a declaraciones de víctimas y en casos de abusos sexuales, casi exclusivamente. El interés por esta técnica, ha llevado a realizar investigaciones a poner a prueba la validez del CBCA en muestras de agresores adultos, llegando a obtener resultados diversos y contradictorios. Con el objetivo de arrojar luz sobre la capacidad del CBCA para discriminar entre declaraciones reales y falsas y validar la Hipótesis Undeutsch en agresores adultos, nos planteamos la elaboración de una revisión meta-analítica de la literatura científica. Los resultados validan la Hipótesis Undeutsch, esto es, las memorias de hechos vividos difieren en contenido y calidad de las memorias de hechos inventados de agresores adultos y, además, el CBCA discrimina entre ambos tipos de memorias. No obstante, no todos los criterios de realidad son aplicables a muestras de agresores, ni son generalizables a otros contextos y muestras. Son necesarios más estudios con rigor científico, elaborados bajo condiciones de *high fidelity*, para alcanzar decisiones más robustas. Se discuten los resultados para su aplicación a la práctica forense.

**Palabras clave:** análisis de contenido; testimonio; agresores; CBCA.

##### Abstract

Content analysis of a statement is a method use to assess credibility statements in judicial context. Concretely, the most utilized assessment tool is Criteria-Based Content Analysis (CBCA), which has been originally applied to victim's statements and in sexual abuse cases, almost exclusively. The interest for this technique, has led to investigate the validity of CBCA in adult offender samples, reaching diverse and contradictory outcomes. With the objective to shed light about the capacity of CBCA to discriminate between real and false accounts and to validate the Undeutsch Hypothesis, we have conducted a meta-analytic review of the scientific literature. Results have validated Undeutsch Hypothesis, that is, memories of self-experienced events differ in content and quality from memories of invented events of adult offenders and CBCA has distinguished between both memories, as well. Notwithstanding, nor every reality criteria are applicable to offender's samples, nor generalizable to other contexts and samples. More research with scientific rigor is needed, made in high fidelity conditions, for reaching more robust decisions. Implications for forensic practice are discussed.

**Keywords:** content analysis; testimony; offenders, CBCA.

## Resumo

A análise de contido da declaración é unha das metodoloxías utilizadas no contexto xudicial para a avaliación da credibilidade dunha testemuña. Concretamente, o instrumento de avaliación máis utilizado é o Criteria-Based Content Analysis (CBCA), que se aplicou orixinalmente a declaración de vítimas en casos de abusos sexuais, case exclusivamente. O interese por esa técnica levou a realizar investigación e a poñer a proba a validez do CBCA en mostrar de agresores adultos, chegando a obterse resultados diversos e contraditorios. Co obxectivo de obter luz sobre a capacidade do CBCA para discriminar entre declaracións reais e falsas e validar a Hipótese Undeutsch en agresores adultos, plantexámonos a elaboración dunha revisión meta-analítica da literatura científica. Os resultados validan a Hipótese Undeutsch, é dicir, as memorias de feitos vividos difiren en contido e calidade das memorias de feitos inventados de agresores adultos e, ademais, o CBCA discrimina entre ambos os tipos de memoria. Non obstante, non todos os criterios de realidade son aplicables á mostra de agresores, nin xeneralizables a outros contextos e mostras. Son necesarios máis estudo con rigor científico, elaborados baixo condicións de *high fidelity*, para alcanzar decisións máis robustas. Discútese os resultados para a súa aplicación á práctica forense.

**Palabras chave:** análise de contido; testemuña; agresores; CBCA.

## Introducción

El análisis de la credibilidad del testimonio se utiliza en el contexto judicial para dotar de valor de prueba al testimonio del denunciante, especialmente al de la víctima. El instrumento más utilizado para ello y con mayor respaldo en la comunidad científica (Amado, Arce, y Fariña, 2015) es el Criteria-Based Content Analysis (Steller y Köhnken, 1989), creado originariamente, para evaluar el testimonio de menores víctimas de abusos sexuales. Aunque el CBCA nace en un contexto y para una población determinados, literatura incipiente ha tratado de generalizar su uso a otras casuísticas (e.g. violencia de género), poblaciones (e.g. adultos) y actores dentro del proceso penal (e.g. testigos, agresores). Si bien en nuestro ordenamiento jurídico se le presuponen al acusado una serie de garantías constitucionales entre las que se encuentra el derecho a mentir, el estudio de la credibilidad del testimonio se presenta como una posibilidad cuando existe una declaración de tamaño suficiente para el análisis de contenido. Para ello, se ha propuesto la aplicación del CBCA, el cual se sustenta en la llamada Hipótesis Undeutsch, esto es, las memorias de hechos experimentados o realmente vividos difieren en contenido y calidad de aquellas que son fruto de la imaginación o fantasía (Undeutsch, 1967).

Son escasas las investigaciones que estudian la validez de la hipótesis y, por extensión, del CBCA en muestras de agresores adultos aunque ya se suponía su aplicabilidad a otras muestras y contextos (Berliner y Conte, 1993) al tratarse de una hipótesis que se sustenta en contenidos de memoria. Por todo lo anterior y como consecuencia de los resultados contradictorios que arroja la literatura científica al respecto, nos planteamos un meta-análisis que someta a contraste la validez de la hipótesis Undeutsch y del CBCA en muestras de agresores de mayoría de edad.



## Método

### *Búsqueda de literatura*

Se ha llevado a cabo una búsqueda bibliográfica en la base de datos científica de referencia (Web of Science) y en redes sociales científicas para el intercambio de información con otros investigadores (i.e., Researchgate). Asimismo, nos hemos puesto en contacto con especialistas en la materia para identificar comunicaciones que no hubieran sido publicadas en revistas. Además, se revisaron las referencias bibliográficas de artículos seleccionados en busca de trabajos no identificados hasta el momento (*ancestry approach*). Las palabras clave utilizadas como motor de búsqueda han sido: análisis de contenido/content analysis, credibilidad/credibility, análisis de contenido basado en criterios/criteria-based content analysis, CBCA, testimonio/testimony, declaración/statement, adultos/adults, agresor/offender.

### *Criterios de inclusión y exclusión*

Para que los trabajos identificados pasaran a formar parte del presente meta-análisis, debían cumplir una serie de criterios de inclusión: 1) que la población de estudio la constituyeran adultos, esto es, con edades iguales o superiores a los 17 años, criterio establecido por los propios autores de las investigaciones primarias para considerar a un sujeto como adulto, 2) que se tratara de una muestra de agresores, 3) que proporcionaran un tamaño del efecto o, en su caso, cualquier otro estadístico que permitiera su cómputo, 4) que analizaran la capacidad discriminativa de los criterios de realidad y del CBCA en su conjunto para diferenciar entre declaraciones reales y falsas.

Se han excluido aquellos estudios en los que la unidad de análisis no era la declaración o que utilizaran otros sistemas categoriales metódicos que no cumplieran el criterio de 'exclusión mutua'. Por otro lado, se han incorporado criterios explícitamente formulados como adicionales al CBCA.

Finalmente, 10 estudios que cumplían los criterios de inclusión, formaron parte del meta-análisis. De ellos se extrajeron 12 tamaños del efecto.

### *Procedimiento*

Una vez identificados los estudios, se procedió a su codificación atendiendo a sus características. Aquellas que pueden tener una influencia en los resultados se tomarán como variables moderadoras, procediéndose al cálculo de un nuevo meta-análisis con base a dichas variables. La jurisprudencia ha establecido que para que una prueba sea considerada evidencia científica admisible en un proceso penal, es necesario que se cumpla el criterio Daubert de publicación en revistas con revisión por pares. Por ello, se tomará este criterio como posible moderador.

La codificación fue llevada a cabo por dos investigadores de forma independiente, obteniendo una concordancia total entre ellos.

### *Análisis de datos*

Se tomaron los tamaños del efecto de los estudios primarios cuando se informaba de ellos explícitamente. En caso de que no se informaran, a partir de las medias y desviaciones típicas o error estándar de la media, se computaron los tamaños del efecto expresados en  $d$  de Cohen. Además, se utilizaron fórmulas de conversión (Rosenthal, 1994) a tamaño del efecto cuando los resultados estaban en  $t$ ,  $F$  o  $\chi^2$ .

El meta-análisis se llevó a cabo siguiendo el procedimiento establecido por Hunter y Schmidt (2015). Se procedió al cálculo del tamaño del efecto ponderado por el tamaño de la muestra ( $d_w$ ) tomando como unidad de análisis del número de declaraciones, corregido posteriormente por una estimación de la fiabilidad del criterio ( $\delta$ ). Además, se calculó el porcentaje de varianza explicada por los errores artificiales (%VE), el intervalo de confianza al 95% (esto es, si el tamaño del efecto obtenido es significativo) y el intervalo de credibilidad al 80% (si los resultados son generalizables a otras muestras y condiciones).

#### *Fiabilidad del criterio*

Se tomaron las fiabilidades inter-codificadores (no así la concordancia entre codificadores por no ser un indicador de fiabilidad) proporcionadas por las investigaciones primarias. Ante la ausencia de coeficientes de fiabilidad en muchos de los estudios incluidos, se procedió a una estimación de la misma a partir de las fiabilidades aportadas para cada uno de los criterios de realidad, obteniéndose así, una fiabilidad para el criterio de .61. Asimismo, se calculó la fiabilidad para el total del CBCA utilizando la fórmula de predicción de Spearman-Brown, obteniéndose un coeficiente de .97.

## **Resultados**

#### *Estudio de outliers*

Previamente a la realización del meta-análisis, se procedió a la búsqueda e identificación de valores extremos, tomando como criterio  $\pm 3 * IQR$  del tamaño del efecto ponderado por el tamaño de la muestra, siempre y cuando no constituyan más del 10% de los tamaños del efecto (Tukey, 1960), puesto que podría tratarse de moderadores y no de valores outliers.

#### *Meta-análisis de la Hipótesis Undeutsch y CBCA en agresores*

Los resultados de los meta-análisis (ver Tabla 1) demuestran un efecto positivo (en dirección a lo apuntado por la Hipótesis Undeutsch) en los criterios de realidad así como en la puntuación total del CBCA. Sin embargo, los tamaños del efecto han resultado ser significativos (el intervalo de confianza no contiene el valor 0) sólo para los criterios 1-7, 9, 13, 15 y 19, así como para el total del instrumento, y generalizables (el intervalo de credibilidad no contiene el valor 0) para los criterios 'admisión de falta de memoria', 'características de la agresión' y para el total del CBCA, esto es, generalizables a otras muestras y contextos.

Los resultados en los meta-análisis para los criterios adicionales muestran un efecto positivo, significativo y generalizable para el criterio 'display insecurities'. Sin embargo, el

efecto encontrado para el criterio “providing reasons for lack of memory” no alcanza la significación.

Finalmente, el promedio de todos los criterios informa de un tamaño del efecto positivo, pequeño (Cohen, 1988), significativo pero no generalizable.

### *Estudio de Moderadores*

Los resultados obtenidos tomando como variable moderadora el criterio Daubert (ver Tabla 2) de revisión por pares, han mostrado tamaños del efecto positivos para los criterios 1-9, 13,15 y 19 así como para la puntuación total del CBCA. Sin embargo, los tamaños del efecto no han sido significativos para los criterios 1, 2, 4, 8 y 19, pero sí generalizables en el caso de los criterios 3, 5 y 15. Los restantes criterios de realidad han mostrado un efecto negativo aunque no significativo.

### **Discusión**

De los resultados anteriores se pueden extraer las siguientes conclusiones. Primera, los resultados validan la hipótesis Undeutsch, esto es, los criterios de realidad discriminan entre memorias de hechos auto-experimentados y hechos fabricados. Segunda, los resultados también validan el CBCA y su capacidad discriminativa entre ambas memorias. No obstante, no todos los criterios de realidad son aplicables a muestras de agresores puesto que no alcanzan la significatividad. Tercero, siguiendo el criterio Daubert para la evidencia científica, los resultados muestran que los criterios cantidad de detalles, descripción de interacciones y admisión de falta de memoria discriminan significativamente entre declaraciones reales y falsas y, a diferencia de lo que ocurría en el meta-análisis global, son generalizables a otros contextos y poblaciones. Sin embargo, sigue habiendo criterios que no pueden utilizarse en muestras de adultos agresores y en cualquier contexto.

Respecto a las implicaciones prácticas de los resultados aquí obtenidos, si bien se valida la Hipótesis Undeutsch y el CBCA, no se puede generalizar su uso a la práctica forense. No todos los criterios son válidos y generalizables a la práctica, son necesarios más estudios de campo y experimentales (*high fidelity*) que pongan a prueba la capacidad discriminativa de los criterios de realidad en el contexto de evaluación determinado, siguiendo un procedimiento “bottom up” (Arce y Fariña, 2009), de concreción de categorías de realidad. Finalmente, es necesario establecer un criterio de decisión estricto, avalado por la evidencia, y no basar la realidad del testimonio en un mero juicio clínico.

Esta revisión meta-analítica está sujeta a las siguientes limitaciones. Primero, es necesario tomar con cautela los resultados, algunos meta-análisis pueden estar sujetos a cierta variabilidad ya que tamaños de muestra inferiores a 400 no garantizan la estabilidad de las estimaciones de muestreo. Segundo, La fiabilidad de criterio es una estimación, dada la ausencia de su cómputo en los estudios primarios. Tercero, se desconoce si determinados estudios siguieron los estándares del SVA que constituyen un requerimiento para la aplicación del CBCA.

Tabla 1. Meta-análisis de los Criterios de Realidad Individuales y la Puntuación Total del CBCA

CRITERIOS	k	N	$d_w$	$SD_d$	$SD_{pre}$	$SD_{res}$	$\delta$	$SD_\delta$	%Var	95% $CI_d$	80% $CV_\delta$
1. Estructura lógica	8	769	0.46	0.6542	0.2242	0.6145	0.60	0.7997	12	0.32, 0.60	-0.42, 1.63
2. Producción inestructurada	6	709	0.73	1.6117	0.2321	1.1382	0.95	1.4812	4	0.57, 0.89	-0.95, 2.84
3. Cantidad de detalles	9	1,025	0.68	1.0479	0.2300	1.0223	0.89	1.3303	5	0.56, 0.80	-0.81, 2.59
4. Engranaje contextual	7	749	0.31	0.7352	0.2031	0.4993	0.40	0.9195	8	0.17, 0.45	-0.78, 1.57
5. Descripción de interacciones	8	1,005	0.62	0.8024	0.2151	0.7731	0.80	1.0060	7	0.50, 0.74	-0.48, 2.09
6. Reproducción conversaciones	11	1,067	0.15	0.5715	0.0000	0.5899	0.19	0.7677	6	0.03, 0.27	-0.79, 1.18
7. Complicaciones inesperadas	8	769	0.25	0.4162	0.2109	0.3588	0.33	0.4669	26	0.11, 0.39	-0.27, 0.93
8. Detalles inusuales	9	809	0.10	0.2299	0.2130	0.0865	0.13	0.1126	86	-0.04, 0.24	-0.01, 0.28
9. Detalles periféricos	8	769	0.16	0.3870	0.2075	0.3266	0.21	0.4251	29	0.02, 0.30	-0.33, 0.75
10. Detalles incomprensivos	2	90	0.17	0.0103	0.3026	0.0000	0.22	0.0000	100	-0.24, 0.58	0.22
11. Asociaciones externas	8	769	0.08	0.4002	0.2057	0.3433	0.11	0.4468	26	-0.06, 0.22	-0.46, 0.68
12. Estado mental subjetivo	9	1,025	0.10	0.5086	0.1894	0.4720	0.14	0.6142	14	-0.02, 0.22	-0.65, 0.92
13. Estado mental autor del delito	6	709	0.22	0.3900	0.1897	0.3407	0.29	0.4433	24	0.08, 0.36	-0.27, 0.86
14. Correcciones espontáneas	8	769	0.07	0.2736	0.2055	0.1806	0.09	0.2350	56	-0.07, 0.21	-0.20, 0.40
15. Admisión falta de memoria	8	722	0.43	0.2785	0.2278	0.1602	0.56	0.2085	67	0.29, 0.57	0.29, 0.83
16. Dudas propio testimonio	3	276	0.02	0.2383	0.2092	0.1140	0.03	0.1483	77	-0.21, 0.25	-0.15, 0.22
17. Autodesaprobación	4	366	0.07	0.4783	0.2102	0.4296	0.09	0.5591	19	-0.13, 0.27	-0.62, 0.81
18. Perdón autor del delito	2	216	0.10	0.0969	0.1940	0.0000	0.13	0.0000	100	-0.17, 0.37	0.13
19. Características agresión	3	298	0.27	0.2147	0.2082	0.0524	0.36	0.0682	94	0.03, 0.50	0.27, 0.44
TOTAL SCORE	6	581	0.31	0.2632	0.2129	0.1547	0.40	0.2013	65	0.15, 0.47	0.14, 0.66
CRITERIOS ADICIONALES											
Display insecurities	2	238	0.46	0.4051	0.1883	0.3587	0.54	0.4178	21	0.20, 0.71	0.00, 1.07
Reasons for lack of memory	2	238	0.21	0.0388	0.1847	0.0000	0.24	0.0000	100	-0.04, 0.46	0.24
PROMEDIO (todos los criterios)	11	1,067	0.27	0.4662	0.2024	0.3743	0.35	0.4975	41	0.15, 0.39	-0.29, 0.99

Nota. k = número de tamaños del efecto; N = tamaño de la muestra;  $d_w$  = tamaño del efecto ponderado por el tamaño de la muestra;  $SD_d$  = desviación estándar de  $d$ ;  $SD_{pre}$  = desviación estándar de  $d$  corregida por artefactos;  $SD_{res}$  = desviación estándar residual;  $\delta$  = tamaño del efecto corregido por la fiabilidad del criterio;  $SD_\delta$  = desviación estándar de  $\delta$ ; %Var = varianza debida a todos los errores artificiales; 95%  $CI_d$  = intervalo de confianza de  $d$  al 95%; 80%  $CV_\delta$  = intervalo de credibilidad de  $\delta$  al 80%

Tabla 2. Meta-análisis del criterio Daubert

CRITERIOS	k	N	d <sub>w</sub>	SD <sub>d</sub>	SD <sub>pre</sub>	SD <sub>res</sub>	δ	SD <sub>δ</sub>	%Var	95% CI <sub>d</sub>	80% CV <sub>δ</sub>
1. Estructura lógica	5	278	0.12	0.5192	0.2709	0.4429	0.16	0.5763	24	-0.11, 0.35	-0.57, 0.90
2. Producción inestructurada	3	218	0.07	0.1189	0.2358	0.0000	0.09	0.0000	100	-0.20, 0.34	0.09
3. Cantidad de detalles	6	534	0.57	0.5071	0.2411	0.4461	0.75	0.5805	23	0.39, 0.75	0.01, 1.49
4. Engranaje contextual	5	258	0.07	0.5277	0.2504	0.4645	0.09	0.6044	22	-0.16, 0.30	-0.68, 0.86
5. Descripción de interacciones	5	514	0.26	0.2786	0.2045	0.1891	0.34	0.2461	54	0.08 0.44	0.03, 0.66
6. Reproducción conversaciones	8	576	0.21	0.4586	-	0.5208	0.46	0.68	29	0.05, 0.37	-0.59, 1.14
7. Complicaciones inesperadas	5	278	0.40	0.5362	0.2822	0.4559	0.53	0.5933	28	0.16, 0.63	-0.23, 1.29
8. Detalles inusuales	6	318	0.21	0.3310	0.2798	0.1749	0.28	0.2277	72	-0.01, 0.42	-0.01, 0.57
9. Detalles periféricos	5	278	0.27	0.6268	0.2755	0.4122	0.36	0.7327	19	0.03, 0.50	-0.58, 1.29
11. Asociaciones externas	5	278	-0.08	0.2591	0.2702	0.0000	-0.10	0.0000	100	-0.31, 0.15	-0.10
12. Estado mental subjetivo	6	534	-0.12	0.2650	0.2143	0.1559	-0.15	0.2029	65	-0.30, 0.06	-0.41, 0.10
13. Estado mental autor delito	3	218	0.35	0.2125	0.2457	0.0000	0.46	0.0000	100	0.07, 0.62	0.46
14. Correcciones espontáneas	5	278	-0.01	0.4296	0.2697	0.3344	-0.01	0.4351	39	-0.24, 0.22	-0.57, 0.55
15. Admisión falta de memoria	7	574	0.44	0.3118	0.2384	0.2010	0.57	0.2615	58	0.28, 0.60	0.24, 0.91
16. Dudas propio testimonio	2	128	-0.14	0.2660	0.2526	0.0835	-0.18	0.1087	90	-0.49, 0.21	-0.32, -0.04
17. Autodesaprobación	3	218	-0.19	0.4641	0.2384	0.3982	-0.24	0.5181	26	-0.46, 0.08	-0.91, 0.42
19. Características agresión	2	150	0.12	0.2134	0.2330	0.0000	0.16	0.0000	100	-0.19, 0.43	0.16
<b>TOTAL SCORE</b>	4	238	0.61	0.0921	0.2666	0.0000	0.62	0.0000	100	0.35, 0.86	0.62

Nota. Los criterios 10 y 18 no pudieron computarse por k insuficientes.

## Referencias

- Amado, B. G., Arce, R., y Fariña, F. (2015). Undeutsch hypothesis and Criteria Based Content Analysis: A meta-analytic review. *European Journal of Psychology Applied to Legal Context*, 7, 3-12.
- Arce, R., y Fariña, F. (2009). Evaluación psicológico forense de la credibilidad y daño psíquico en casos de violencia de género mediante el Sistema de Evaluación Global. En F. Fariña, R. Arce, y G. Buela-Casal (Eds.), *Violencia de género. Tratado psicológico y legal*. Madrid, España: Biblioteca Nueva.
- Berliner, L., y Conte, J. R. (1993). Sexual abuse evaluation: Conceptual and empirical obstacles. *Child Abuse and Neglect*, 17, 111-125.
- \*Caso, L., Vrij, A., Mann, S., y de Leo, G. (2006). Deceptive responses: The impact of verbal and non-verbal countermeasures. *Legal and Criminological Psychology*, 11, 99-111.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2a. ed.). Hillsdale, NJ: LEA.
- \*Dana-Kirby, L. (1997). *Discerning truth from deception: Is Criteria-Based Content Analysis effective with adult statements?* (Tesis doctoral inédita). University of Oregon, Eugene.
- \*Gödert, H. W., Gamer, M., Rill, H. G., y Vossel, G. (2005). Statement Validity Assessment: Inter-rater reliability of Criteria-Based Content Analysis in the mock-crime paradigm. *Legal and Criminological Psychology*, 10, 225-245.
- Hunter, J. E., y Schmidt, F. L. (2015). *Methods of meta-analysis: Correcting error and bias in research findings* (3a. ed.). Newbury Park, CA: Sage.
- \*Johnston, S., Candelier, A., Powers-Green, D., y Rahmani, S. (2014). Attributes of truthful versus deceitful statements in the evaluation of accused child molesters. *Sage Open*, 4(3), 1-10.
- \*Lee, Z., Klaver, J. R., y Hart, S. D. (2008). Psychopathy and verbal indicators of deception in offenders. *Psychology, Crime y Law*, 14, 73-84.
- \*Porter, S., y Yuille, J. C. (1996). The language of deceit: An investigation of the verbal clues to deception in the interrogation context. *Law and Human Behavior*, 20, 443-458.
- Rosenthal, R. (1994). Parametric measures of effect size. En H. Cooper y L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 231-244). New York, NY: Russell Sage Foundation.
- Steller, M., y Köhnken, G. (1989). Criteria-Based Content Analysis. En D.C. Raskin (Ed.), *Psychological methods in criminal investigation and evidence* (pp. 217-245). New York, NY: Springer-Verlag.
- \*Ternes, M. (2009). *Verbal credibility assessment of incarcerated violent offenders' memory reports*. (Tesis doctoral inédita). University of British Columbia, Vancouver.
- Tukey, J. W. (1960). A survey of sampling from contaminated distributions. En I. Olkin, J.G. Ghurye, W. Hoeffding, W.G. Madow, y H. Mann (Eds.), *Contributions to probability and statistics* (pp. 448-485). Stanford, CA: Stanford University Press.
- Undeutsch, U. (1967). Beurteilung der glaubhaftigkeit von aussagen. En U. Undeutsch (Ed.), *Handbuch der psychologie, Vol. 11: Forensische psychologie* (pp. 26-181). Göttingen, Germany: Hogrefe.
- \*Vrij, A., y Mann, S., (2006). Criteria-Based Content Analysis: An empirical test of its underlying processes. *Psychology, Crime and Law*, 12, 337-349.
- \*Vrij, A., Mann, S., Kristen, S., y Fisher, R. P. (2007). Cues to deception and ability to detect lies as a function of police interview styles. *Law and Human Behavior*, 31, 449-518.
- \*Willén, R. M., y Strömwall, L. A. (2011). Offender's uncoerced false confessions: A new application of statement analysis? *Legal and Criminological Psychology*, 17, 346-359.

## 9. ÍNDICE DE TABLAS<sup>3</sup>



---

<sup>3</sup> Solamente en la primera de las tablas se recoge, en una nota al pie, las definiciones de cada uno de los índices de los que consta. En las restantes tablas se han obviado por tratarse de los mismos.





Table 3. Results of Global Meta-Analysis for Individual Reality Criteria and Total CBCA Score, and Additional Criteria

CBCA Criterion	k	n	$d_w$	$SD_d$	$SD_{pre}$	$SD_{res}$	$\delta$	$SD_\delta$	%Var	95% $CI_d$	80% $CV_\delta$
1. Logical structure	30	2,265	0.48	0.6990	0.2503	0.6527	0.62	0.8493	13	0.40, 0.56	-0.46, 1.71
2. Unstructured production	27	1,987	0.53	0.9241	0.2570	0.8876	0.69	1.1551	8	0.45, 0.61	-0.79, 2.17
3. Quantity of details	35	2,714	0.55	0.8294	0.2529	0.7899	0.71	1.0279	9	0.47, 0.63	-0.60, 2.03
4. Contextual embedding	29	2,137	0.19	0.6169	0.2372	0.5868	0.24	0.7411	15	0.11, 0.27	-0.70, 1.19
5. Description of interactions	29	2,243	0.27	0.3742	0.2349	0.2912	0.36	0.3790	39	0.19, 0.35	-0.13, 0.84
6. Reproduction conversations	34	2,528	0.34	0.4990	0.1780	0.4662	0.44	0.6067	13	0.26, 0.42	-0.33, 1.22
7. Unexpected complications	29	1,956	0.25	0.3788	0.2498	0.2847	0.32	0.3705	43	0.17, 0.33	-0.15, 0.79
8. Unusual details	35	2,441	0.31	0.6532	0.2489	0.6039	0.41	0.7859	14	0.23, 0.39	-0.59, 1.42
9. Superfluous details	27	1,863	0.14	0.5676	0.2437	0.5126	0.18	0.6670	18	0.04, 0.24	-0.67, 1.04
10. Details misunderstood	5	376	0.22	0.1208	0.2357	0.0000	0.28	0.0000	100	0.02, 0.42	0.28
11. External associations	22	1,612	0.26	0.4781	0.2405	0.3268	0.34	0.5376	25	0.16, 0.36	-0.35, 1.02
12. Subjective mental state	28	2,170	0.18	0.4843	0.2312	0.4256	0.23	0.5538	23	0.10, 0.26	-0.47, 0.94
13. Perpetrator's mental state	31	2,232	0.09	0.6212	0.2376	0.5741	0.11	0.7470	15	0.01, 0.17	-0.84, 1.07
14. Spontaneous corrections	29	1,842	0.16	0.5276	0.2545	0.4622	0.20	0.6014	23	0.06, 0.26	-0.56, 0.97
15. Admitting lack of memory	34	2,305	0.25	0.3823	0.2494	0.2897	0.32	0.3770	42	0.17, 0.33	-0.16, 0.80
16. Doubts one's testimony	26	1,755	0.20	0.4521	0.2478	0.3781	0.26	0.4919	30	0.10, 0.30	-0.37, 0.89
17. Self-deprecation	13	948	0.04	0.4629	0.2354	0.3985	0.05	0.5186	26	-0.08, 0.16	-0.61, 0.71
18. Pardoning the perpetrator	8	680	-0.02	0.2796	0.2178	0.1753	-0.02	0.2281	61	-0.18, 0.14	-0.31, 0.27
19. Details characteristics offence	5	562	0.28	0.1894	0.1966	0.0000	0.36	0.0000	100	0.12, 0.44	0.36
Total CBCA Score	31	2,124	0.55	0.6759	0.2475	0.6290	0.56	0.6386	13	0.47, 0.63	-0.25, 1.37
Average (original criteria)	46	3,223	0.25	0.5032	0.2368	0.4269	0.33	0.5614	32	0.17, 0.33	-0.39, 1.05
Additional Criteria											
Reporting style	3	357	0.41	0.2030	0.1874	0.0781	0.48	0.0909	85	0.20, 0.63	0.36, 0.59
Display insecurities	3	297	0.67	0.5540	0.2111	0.5122	0.78	0.5965	14	0.43, 0.90	0.01, 1.54
Providing reasons lack memory	4	447	0.15	0.2877	0.1902	0.2158	0.18	0.2514	44	-0.03 0.33	-0.14, 0.50
Clichés	3	267	-0.18	0.5145	0.2134	0.4682	-0.21	0.5452	17	-0.41, 0.05	-0.90, 0.49
Repetitions	4	417	-0.47	0.5851	0.2011	0.5494	-0.54	0.6399	12	-0.67, -0.27	-1.36, 0.27
Average (original + additional)	46	3,223	0.27	0.4821	0.2313	0.4053	0.34	0.5275	34	0.19, 0.35	-0.33, 1.01

Note.  $k$  = number of studies;  $n$  = total sample size;  $d_w$  = effect size weighted for sample size;  $SD_d$  = observed standard deviation of  $d$ ;  $SD_{pre}$  = standard deviations of observed  $d$ -values corrected from all artifacts;  $SD_{res}$  = standard deviation of observed  $d$ -values after removal of variance due to all artifacts;  $\delta$  = effect size corrected for criterion reliability;  $SD_\delta$  = standard deviation of  $\delta$ ; %Var = variance accounted for by artefactual errors; 95%  $CI_d$  = 95% confidence interval for  $d$ ; 80%  $CV_\delta$  = 80% credibility interval for  $\delta$ .

Table 4. Results of the Meta-analysis of Daubert Standard Peer Review Moderator Variable

CRITERIA	k	N	$d_w$	$SD_d$	$SD_{pre}$	$SD_{res}$	$\delta$	$SD_\delta$	% Var	95% $CI_d$	90% $CV_\delta$
1. Logical structure	21	1,376	0.44	0.6489	0.2636	0.5929	0.57	0.7715	16	0.34, 0.54	-0.41, 1.56
2. Unstructured production	18	1,152	0.20	0.6019	0.2547	0.5453	0.26	0.7097	18	0.08, 0.32	-0.64, 1.17
3. Quantity of details	27	1,905	0.48	0.5982	0.2584	0.5396	0.63	0.7022	19	0.38, 0.58	-0.26, 1.53
4. Contextual embedding	21	1,328	0.06	0.4931	0.2531	0.4231	0.08	0.5506	26	-0.04, 0.16	-0.62, 0.79
5. Description of interactions	22	1,582	0.25	0.3812	0.2422	0.2943	0.32	0.3830	40	0.15, 0.35	-0.16, 0.81
6. Reproduction conversations	25	1,693	0.37	0.4007	0.1557	0.3693	0.48	0.4805	15	0.27, 0.47	-0.14, 1.09
7. Unexpected complications	20	1,121	0.24	0.4439	0.2732	0.3499	0.31	0.4553	38	0.12, 0.36	-0.26, 0.90
8. Unusual details	26	1,606	0.36	0.7070	0.2659	0.6551	0.46	0.8525	14	0.26, 0.46	-0.63, 1.56
9. Superfluous details	19	1,054	0.06	0.5804	0.2703	0.5136	0.07	0.6700	22	-0.06, 0.18	-0.78, 0.93
10. Details misunderstood	4	354	0.22	0.1235	0.2176	0.0000	0.29	0.0000	100	0.004, 0.43	0.29
11. External associations	15	855	0.16	0.3141	0.2685	0.1631	0.21	0.2122	73	0.02, 0.30	-0.06, 0.48
12. Subjective mental state	19	1,335	0.09	0.3713	0.2407	0.2827	0.12	0.3679	42	-0.02, 0.20	-0.34, 0.60
13. Perpetrator's mental state	22	1,397	0.06	0.7077	0.2527	0.6611	0.08	0.8603	13	-0.04, 0.16	-1.01, 1.19
14. Spontaneous corrections	22	1,350	0.12	0.5822	0.2580	0.5219	0.15	0.6792	20	0.02, 0.22	-0.72, 1.02
15 Admitting lack of memory	27	1,813	0.25	0.4184	0.2506	0.3350	0.32	0.4360	36	0.15, 0.35	-0.24, 0.88
16. Doubts one's testimony	20	1,246	0.16	0.4872	0.2569	0.4139	0.21	0.5386	28	0.04, 0.28	-0.47, 0.90
17. Self-deprecation	8	561	-0.13	0.4499	0.2413	0.3798	-0.17	0.4942	29	-0.29, 0.03	-0.80, 0.46
18. Pardoning the perpetrator	4	345	-0.24	0.1989	0.2212	0.0000	-0.31	0.0000	100	-0.45, -0.02	-0.31
19. Details characteristic offence	3	309	0.27	0.2058	0.2045	0.0237	0.35	0.0309	99	0.05, 0.48	0.31, 0.39
Reporting Style	2	209	0.42	0.2610	0.2004	0.1672	0.51	0.1947	59	0.14, 0.69	0.26, 0.76
Display insecurities	2	149	0.56	0.7675	0.2395	0.7291	0.65	0.8491	10	0.23, 0.89	-0.43, 1.74
Providing reasons lack memory	3	299	0.11	0.3437	0.2012	0.2787	0.13	0.3246	34	-0.12, 0.34	-0.29, 0.54
Average	36	2,428	0.20	0.4575	0.2407	0.3733	0.26	0.4786	39	0.12, 0.28	-0.35, 0.87
Total CBCA Score	25	1,493	0.55	.4087	.2650	.3111	0.57	.3261	42	0.45, 0.65	0.15, 0.99

Table 5. Results of the Meta-analysis of Self-Experienced Events Moderator Variable

CRITERIA	k	N	$d_w$	$SD_d$	$SD_{pre}$	$SD_{res}$	$\delta$	$SD_\delta$	%Var	95% CI $_d$	80% CV $_d$
1. Logical structure	23	1,536	0.58	0.8130	0.2720	0.7661	0.75	0.9970	11	0.48, 0.68	-0.52, 2.03
2. Unstructured production	20	1,258	0.69	1.0856	0.2892	1.0463	0.89	1.3616	7	0.57, 0.81	-0.85, 2.64
3. Quantity of details	26	1,852	0.60	0.8655	0.2673	0.8232	0.79	1.0712	9	0.50, 0.70	-0.58, 2.16
4. Contextual embedding	22	1,394	0.22	0.7277	0.2566	0.6809	0.29	0.8861	12	0.12, 0.32	-0.84, 1.42
5. Description of interactions	23	1,573	0.24	0.3864	0.2479	0.2963	0.31	0.3856	41	0.14, 0.34	-0.18, 0.81
6. Reproduction conversations	28	1,821	0.44	0.4573	0.1781	0.4212	0.57	0.5481	15	0.34, 0.54	-0.13, 1.27
7. Unexpected complications	22	1,375	0.31	0.3695	0.2619	0.2607	0.40	0.3392	50	0.21, 0.41	-0.03, 0.83
8. Unusual details	25	1,535	0.18	0.4702	0.2593	0.3922	0.24	0.5104	30	0.08, 0.28	-0.41, 0.89
9. Superfluous details	22	1,386	0.13	0.5455	0.2548	0.4824	0.17	0.6277	22	0.03, 0.23	-0.63, 0.98
10. Details misunderstood	4	317	0.19	0.1114	0.2289	0.0000	0.24	0.0000	100	-0.02, 0.40	0.24
11. External associations	20	1,254	0.37	0.4886	0.2648	0.4107	0.48	0.5344	29	0.25, 0.49	-0.20, 1.16
12. Subjective mental state	26	1,812	0.22	0.5173	0.2451	0.4555	0.29	0.5928	22	0.12, 0.32	-0.46, 1.05
13. Perpetrator's mental state	23	1,430	0.07	0.4931	0.2555	0.4217	0.10	0.5487	27	-0.03, 0.17	-0.60, 0.80
14. Spontaneous corrections	23	1,383	0.10	0.5181	0.2604	0.4479	0.13	0.5829	25	0.002, 0.20	-0.61, 0.88
15. Admitting lack of memory	27	1,873	0.23	0.3922	0.2460	0.3055	0.30	0.3975	39	0.13, 0.33	-0.21, 0.81
16. Doubts one's testimony	17	1,120	0.05	0.4137	0.2479	0.3312	0.06	0.4310	36	-0.07, 0.17	-0.49, 0.61
17. Self-deprecation	13	948	0.04	0.4628	0.2354	0.3985	0.05	0.5186	26	-0.08, 0.16	-0.61, 0.71
18. Pardoning the perpetrator	8	680	-0.02	0.2796	0.2178	0.1753	-0.02	0.2281	61	-0.18, 0.14	-0.31, 0.27
19. Details characteristic offence	5	562	0.28	0.1894	0.1966	0.0000	0.36	0.0000	100	0.12, 0.44	0.36
Reporting Style	2	298	0.32	0.0447	0.1663	0.0000	0.38	0.0000	100	0.08, 0.55	0.3800
Display insecurities	2	238	0.46	0.4051	0.1883	0.3587	0.53	0.4178	21	0.20, 0.71	0.001, 1.07
Providing reasons lack memory	3	388	0.06	0.1863	0.1763	0.0604	0.07	0.0704	89	-0.14, 0.26	-0.02, 0.16
Average	34	2,277	0.26	0.4647	0.2371	0.3879	0.33	0.5022	35	0.22, 0.44	0.01, 0.65
Total CBCA Score	27	1,613	0.55	0.6696	0.2651	0.6148	0.55	0.6243	16	0.45, 0.65	-0.24, 1.35

Tabla 6. Results of the Meta-analysis of Non Self-experienced Events (Witness) Moderator Variable

CRITERIA	k	N	$d_w$	$SD_d$	$SD_{pre}$	$SD_{res}$	$\delta$	$SD_\delta$	%Var	95% $CI_d$	80% $CV_\delta$
1. Logical structure	5	328	0.21	0.3281	0.2518	0.2102	0.27	0.2736	59	-0.01, 0.42	-0.08, 0.62
2. Unstructured production	6	408	0.31	0.5540	0.2519	0.4934	0.41	0.6421	21	0.11, 0.51	-0.41, 1.23
3. Quantity of details	8	541	0.76	0.7422	0.2881	0.6841	0.99	0.8902	15	0.58, 0.94	-0.15, 2.12
4. Contextual embedding	5	342	0.23	0.3276	0.2475	0.2146	0.30	0.2793	57	0.01, 0.44	-0.06, 0.65
5. Description of interactions	5	349	0.36	0.4592	0.2513	0.3843	0.47	0.5000	30	0.14, 0.57	-0.17, 1.11
6. Reproduction conversations	5	386	0.50	0.3643	0.2492	0.2657	0.65	0.3457	47	0.30, 0.70	0.21, 1.09
7. Unexpected complications	6	260	0.13	0.5381	0.3069	0.4420	0.17	0.5752	32	-0.10, 0.36	-0.56, 0.90
8. Unusual details	9	585	0.84	0.9088	0.3014	0.8574	1.09	1.1157	11	0.66, 1.02	-0.33, 2.52
9. Superfluous details	4	156	0.46	0.9415	0.3370	0.8794	0.59	1.1440	13	0.15, 0.77	-0.87, 2.06
13. Perpetrator's mental state	7	481	0.27	1.0007	0.2483	0.9694	0.35	1.2615	6	0.09, 0.45	-1.27, 1.96
14. Spontaneous corrections	7	481	0.31	0.5110	0.2505	0.4454	0.41	0.5796	24	0.13, 0.49	-0.33, 1.15
15 Admitting lack of memory	7	432	0.32	0.3267	0.2640	0.1925	0.41	0.2504	65	0.12, 0.52	0.09, 0.73
16. Doubts one's testimony	7	575	0.49	0.3761	0.2421	0.2877	0.64	0.3745	41	0.33, 0.65	0.16, 1.12
Average	11	745	0.39	0.5835	0.2707	0.5032	0.51	0.6548	65	0.23, 0.55	-0.33, 1.35
Total CBCA Score	5	250	1.16	0.6485	0.3075	0.5710	1.21	0.5986	22	0.88, 1.43	0.45, 1.98

Table 7. Results of the Meta-analysis of Offenders Moderator Variable

CRITERIA	k	N	$d_w$	$SD_d$	$SD_{pre}$	$SD_{res}$	$\delta$	$SD_\delta$	%Var	95% $CI_d$	80% $CV_\delta$
1. Logical structure	8	769	0.46	0.6542	0.2242	0.6145	0.60	0.7997	12	0.32, 0.60	-0.42, 1.63
2. Unstructured production	6	709	0.73	1.6117	0.2321	1.1382	0.95	1.4812	4	0.57, 0.89	-0.95, 2.84
3. Quantity of details	9	1,025	0.68	1.0479	0.2300	1.0223	0.89	1.3303	5	0.56, 0.80	-0.81, 2.59
4. Contextual embedding	7	749	0.31	0.7352	0.2031	0.4993	0.40	0.9195	8	0.17, 0.45	-0.78, 1.57
5. Description of interactions	8	1,005	0.62	0.8024	0.2151	0.7731	0.80	1.0060	7	0.50, 0.74	-0.48, 2.09
6. Reproduction conversations	11	1,067	0.15	0.5715	0.0000	0.5899	0.19	0.7677	-6	0.03, 0.27	-0.79, 1.18
7. Unexpected complications	8	769	0.25	0.4162	0.2109	0.3588	0.33	0.4669	26	0.11, 0.39	-0.27, 0.93
8. Unusual details	9	809	0.10	0.2299	0.2130	0.0865	0.13	0.1126	86	-0.04, 0.24	-0.01, 0.28
9. Superfluous details	8	769	0.16	0.3870	0.2075	0.3266	0.21	0.4251	29	0.02, 0.30	-0.33, 0.75
10. Details misunderstood	2	90	0.17	0.0103	0.3026	0.0000	0.22	0.0000	100	-0.25, 0.59	0.22
11. External associations	8	769	0.08	0.4002	0.2057	0.3433	0.11	0.4468	26	-0.06, 0.22	-0.46, 0.68
12. Subjective mental state	9	1,025	0.10	0.5086	0.1894	0.4720	0.14	0.6142	14	-0.02, 0.22	-0.65, 0.92
13. Perpetrator's mental state	6	709	0.22	0.3900	0.1897	0.3407	0.29	0.4433	24	0.08, 0.36	-0.27, 0.86
14. Spontaneous corrections	8	769	0.07	0.2736	0.2055	0.1806	0.09	0.2350	56	-0.07, 0.21	-0.20, 0.40
15 Admitting lack of memory	8	722	0.43	0.2785	0.2278	0.1602	0.56	0.2085	67	0.29, 0.57	0.29, 0.83
16. Doubts one's testimony	3	276	0.02	0.2383	0.2092	0.1140	0.03	0.1483	77	-0.21, 0.25	-0.15, 0.22
17. Self-deprecation	4	366	0.07	0.4783	0.2102	0.4296	0.09	0.5591	19	-0.13, 0.27	-0.62, 0.81
18. Pardoning the perpetrator	2	216	0.10	0.0969	0.1940	0.0000	0.13	0.0000	100	-0.17, 0.37	0.13
19. Details characteristic offence	3	298	0.27	0.2147	0.2082	0.0524	0.36	0.0682	94	0.03, 0.50	0.27, 0.44
Display insecurities	2	238	0.46	0.4051	0.1883	0.3587	0.54	0.4178	21	0.20, 0.71	0.001, 1.07
Providing reasons lack memory	2	238	0.21	0.0388	0.1847	0.0000	0.24	0.0000	100	-0.04, 0.46	0.24
Average (all criteria)	11	1,067	0.27	0.4662	0.2024	0.3743	0.35	0.4975	41	0.15, 0.39	-0.29, 0.99
Total CBCA Score	6	581	0.31	0.2632	0.2129	0.1547	0.40	0.2013	65	0.15, 0.47	0.14, 0.66

Table 8. Results of the Meta-analysis of Victims Moderator Variable

CRITERIA	k	N	$d_w$	$SD_d$	$SD_{pre}$	$SD_{res}$	$\delta$	$SD_\delta$	%Var	95% CI <sub>d</sub>	80% CV <sub>δ</sub>
1. Logical structure	8	708	0.64	0.8219	0.2477	0.7837	0.83	1.0197	9	0.49. 0.79	-0.47. 2.14
2. Unstructured production	7	490	0.45	0.9601	0.2568	0.9251	0.59	1.2039	7	0.27. 0.63	-0.95. 2.13
3. Quantity of details	8	708	0.24	0.5332	0.2187	0.4863	0.32	0.6328	17	0.09. 0.39	-0.49. 1.13
4. Contextual embedding	7	534	0.19	0.53	0.2330	0.4794	0.24	0.6239	19	0.01. 0.37	-0.55. 1.04
5. Description of interactions	7	534	0.39	0.3675	0.2430	0.2757	0.51	0.3588	44	0.21. 0.56	0.05. 0.97
6. Reproduction conversations	8	560	0.50	0.3359	0.2606	0.2120	0.66	0.2759	60	0.33. 0.67	0.30. 1.01
7. Unexpected complications	6	464	0.38	0.2746	0.2403	0.1328	0.49	0.1728	77	0.19. 0.56	0.27. 0.71
8. Unusual details	8	584	0.15	0.4161	0.2371	0.3419	0.20	0.4449	32	-0.007. 0.31	-0.37. 0.77
9. Superfluous details	7	558	0.09	0.4971	0.2255	0.4430	0.11	0.5765	20	-0.07. 0.05	-0.62. 0.85
11. External associations	4	370	0.66	0.5301	0.2455	0.4698	0.86	0.6114	21	0.45. 0.87	0.07. 1.64
12. Subjective mental state	8	560	0.23	0.4372	0.2447	0.3623	0.30	0.4715	31	0.06. 0.40	-0.30. 0.90
13. Perpetrator's mental state	9	610	0.09	0.4083	0.2448	0.3268	0.12	0.4253	36	-0.07. 0.25	-0.43. 0.66
14. Spontaneous corrections	6	464	0.19	0.7216	0.2317	0.6833	0.25	0.8892	10	0.01. 0.37	-0.88. 1.39
15. Admitting lack of memory	9	700	0.14	0.3237	0.2295	0.2283	0.18	0.2971	50	-0.003. 0.28	-0.20. 0.56
16. Doubts one's testimony	6	464	0.16	0.5335	0.2307	0.4810	0.21	0.6260	19	-0.02. 0.34	-0.59. 1.01
17. Self-deprecation	6	422	0.10	0.4494	0.2407	0.3795	0.13	0.4938	29	-0.09. 0.29	-0.49. 0.77
18. Pardoning the perpetrator	5	396	-0.03	0.3088	0.2259	0.2106	-0.04	0.2741	53	-0.23. 0.17	-0.39. 0.30
19. Details characteristic offence	2	264	0.28	0.1560	0.1826	0.0000	0.37	0.0000	100	0.04. 0.52	0.37
Average (all criteria)	11	840	0.27	0.4781	0.2355	0.4012	0.35	0.5221	35	0.13. 0.41	-0.32. 1.02
Total CBCA Score	10	678	0.70	0.8759	0.2518	0.8389	0.71	0.8518	8	0.54. 0.86	-0.37. 1.80

Table 9. Results of the Meta-analysis of Field Studies Moderator Variable

CRITERIA	k	N	$d_w$	$SD_d$	$SD_{pre}$	$SD_{res}$	$\delta$	$SD_\delta$	%Var	95% CI <sub>d</sub>	80% CV <sub>g</sub>
1. Logical structure	4	370	1.26	0.5516	0.3241	0.4463	1.65	0.5808	34	1.04, 1.47	0.90, 2.39
2. Unstructured production	6	422	0.51	1.0239	0.2604	0.9902	0.66	1.2885	6	0.31, 0.71	-0.98, 2.31
3. Quantity of details	4	370	0.53	0.5992	0.2332	0.5520	0.69	0.7183	15	0.31, 0.74	-0.23, 1.61
4. Contextual embedding	4	346	0.20	0.6488	0.2194	0.6105	0.26	0.7945	11	-0.01, 0.41	-0.76, 1.27
5. Description of interactions	4	346	0.47	0.4260	0.2349	0.3554	0.61	0.4624	30	0.25, 0.68	0.02, 1.20
6. Reproduction conversations	5	372	0.66	0.2747	0.2677	0.0616	0.86	0.0802	95	0.44, 0.87	0.75, 0.96
7. Unexpected complications	5	396	0.35	0.2925	0.2365	0.1721	0.46	0.2239	65	0.15, 0.55	0.17, 0.75
8. Unusual details	5	396	0.28	0.4032	0.2325	0.3295	0.36	0.4288	33	0.08, 0.48	-0.18, 0.91
9. Superfluous details	4	370	0.16	0.5819	0.2111	0.5423	0.21	0.7057	13	-0.04, 0.36	-0.69, 1.12
11. External associations	4	370	0.66	0.5301	0.2455	0.4698	0.86	0.6114	21	0.44, 0.87	0.08, 1.64
12. Subjective mental state	5	372	0.31	0.5168	0.2410	0.4572	0.40	0.5949	22	0.11, 0.51	-0.36, 1.16
13. Perpetrator's mental state	6	422	0.00	0.3828	0.2397	0.2984	0.00	0.3883	39	-0.20, 0.20	-0.50, 0.50
14. Spontaneous corrections	5	396	0.19	0.7810	0.2290	0.7467	0.25	0.9717	8	-0.006, 0.39	-0.99, 1.49
15. Admitting lack of memory	6	422	0.08	0.4044	0.2404	0.3252	0.11	0.4232	35	-0.12, 0.28	-0.43, 0.65
16. Doubts one's testimony	5	396	0.18	0.5752	0.2287	0.5278	0.24	0.6869	16	-0.02, 0.38	-0.64, 1.12
17. Self-deprecation	6	422	0.10	0.4494	0.2407	0.3795	0.13	0.4938	29	-0.09, 0.29	-0.50, 0.77
18. Pardoning the perpetrator	5	396	-0.03	0.3088	0.2259	0.2106	-0.04	0.2741	53	-0.23, 0.17	-0.40, 0.30
19. Details characteristic offence	2	264	0.28	0.1560	0.1826	0.0000	0.37	0.0000	100	0.04, 0.51	0.37
Average (all criteria.18)	6	422	0.34	0.4948	0.2385	0.4153	0.45	0.5404	35	0.14, 0.54	-0.24, 1.14
Average (significant criteria. 10)	6	422	0.53	0.4774	0.2458	0.3834	0.69	0.4989	42	0.33, 0.73	0.05, 1.33
Total CBCA Score	6	288	1.43	1.3109	0.3267	1.2695	1.46	1.2890	6	1.17, 1.69	-0.19, 3.11

Note. <sup>a</sup>Significant criteria (CBCA criteria. as for additional criteria. studies were insufficient): 1-3, 5-8, 11, 12 and 19.

Table 10. Results of the Meta-analysis of Sexual and Intimate Partner Violence (IPV) Field Studies Moderator Variable

CRITERIA	k	N	$d_w$	$SD_d$	$SD_{pre}$	$SD_{res}$	$\delta$	$SD_\delta$	%Var	95% CI <sub>d</sub>	80% CV <sub>δ</sub>
1. Logical structure	3	211	1.14	0.7047	0.3309	0.6222	1.48	0.6557	22	0.85, 1.43	0.44, 2.52
2. Unstructured production	5	263	1.23	0.5473	0.3759	0.6131	1.60	0.5176	47	0.97, 1.48	0.94, 2.26
3. Quantity of details	3	211	1.04	0.1310	0.1011	0.0000	1.36	0.0000	100	0.75, 1.33	1.36
4. Contextual embedding	3	187	0.56	0.7062	0.2791	0.6587	0.73	0.8441	16	0.27, 0.85	-0.34, 1.81
5. Description of interactions	3	187	0.59	0.5524	0.2811	0.4755	0.76	0.6188	26	0.29, 0.88	-0.03, 1.55
6. Reproduction conversations	4	213	0.86	0.1896	0.3281	0.0000	1.12	0.0000	100	0.58, 1.13	1.12
7. Unexpected complications	4	237	0.52	0.2736	0.2820	0.0000	0.68	0.0000	100	0.26, 0.77	0.68
8. Unusual details	4	237	0.55	0.3053	0.2840	0.1121	0.71	0.1459	86	0.29, 0.80	0.52, 0.90
9. Superfluous details	3	211	0.65	0.2159	0.2726	0.0000	0.85	0.0000	100	0.37, 0.92	0.85
11. External associations	3	211	1.08	0.2776	0.3231	0.0000	1.41	0.0000	100	0.79, 1.37	1.41
12. Subjective mental state	4	213	0.67	0.3984	0.3084	0.2523	0.87	0.3281	60	0.39, 0.94	0.45, 1.29
13. Perpetrator's mental state	5	263	0.11	0.4514	0.2784	0.1262	0.14	0.4623	38	-0.12, 0.34	-0.44, 0.73
14. Spontaneous corrections	5	237	0.75	0.4795	0.3032	0.3716	0.98	0.4835	40	0.49, 1.00	0.36, 1.60
15 Admitting lack of memory	5	263	0.36	0.2330	0.2875	0.0000	0.47	0.0000	100	0.12, 0.59	0.47
16. Doubts one's testimony	4	237	0.61	0.3047	0.2895	0.0949	0.80	0.1235	90	0.35, 0.86	0.64, 0.95
17. Self-deprecation	5	263	0.44	0.1611	0.2918	0.0000	0.57	0.0000	100	0.20, 0.67	0.57
18. Pardoning the perpetrator	4	237	0.19	0.1661	0.2642	0.0000	0.25	0.0000	100	-0.06, 0.44	0.25
Average (all criteria. 17) <sup>a</sup>	5	263	0.67	0.3587	0.2871	0.1957	0.87	0.2459	72	0.41, 0.92	0.55, 1.18
Average (significant criteria. 15) <sup>b</sup>	5	263	0.74	0.3654	0.2892	0.2134	0.96	0.2478	72	0.48, 0.99	0.64, 1.28
Total CBCA Score	5	209	2.05	0.9975	0.3860	0.9198	2.15	0.9642	15	1.72, 2.38	0.91, 3.38

Note. Meta-analysis for 10 and 19 criteria were not computed due to insufficient k.

<sup>a</sup> Significant criteria (CBCA criteria): 1-9, 11-18.

<sup>b</sup> Significant criteria (CBCA criteria): 1-9, 11, 12, 4-17.



Table 11. Results of the Meta-analysis of Experimental Studies Moderator Variable

CRITERIA	k	N	d <sub>w</sub>	SD <sub>d</sub>	SD <sub>pre</sub>	SD <sub>res</sub>	δ	SD <sub>δ</sub>	%Var	95% CI <sub>d</sub>	80% CV <sub>δ</sub>
1. Logical structure	25	1,815	0.31	0.6268	0.2441	0.5773	0.41	0.7512	15	0.21, 0.41	-0.55, 1.37
2. Unstructured production	21	1,565	0.53	0.8952	0.2561	0.8578	0.69	1.1163	8	0.59, 0.79	-0.73, 2.12
3. Quantity of details	31	2,344	0.55	0.8601	0.2559	0.8211	0.71	1.0686	9	0.47, 0.63	-0.65, 2.08
4. Contextual embedding	24	1,711	0.17	0.6192	0.2405	0.5705	0.22	0.7424	15	0.07, 0.27	-0.73, 1.17
5. Description of interactions	25	1,897	0.24	0.3526	0.2356	0.2623	0.31	0.3413	45	0.10, 0.38	-0.12, 0.75
6. Reproduction conversations	29	2,156	0.29	0.5085	0.1604	0.4825	0.37	0.6279	10	0.21, 0.37	-0.43, 1.18
7. Unexpected complications	24	1,560	0.22	0.3929	0.2534	0.3003	0.28	0.3908	41	0.12, 0.32	-0.21, 0.78
8. Unusual details	30	2,045	0.32	0.6911	0.2520	0.6435	0.42	0.8373	13	0.24, 0.40	-0.65, 1.49
9. Superfluous details	23	1,493	0.13	0.5638	0.2512	0.5048	0.17	0.6569	20	0.03, 0.23	-0.66, 1.02
10. Details misunderstood	4	217	0.29	0.1039	0.2797	0.0000	0.38	0.0000	100	0.01, 0.56	0.38
11. External associations	18	1,242	0.14	0.3881	0.2438	0.3019	0.18	0.3929	39	0.02, 0.26	-0.32, 0.68
12. Subjective mental state	23	1,798	0.15	0.4730	0.2294	0.4136	0.20	0.5382	23	0.05, 0.25	-0.49, 0.89
13. Perpetrator's mental state	25	1,810	0.10	0.6631	0.2372	0.6192	0.14	0.5058	13	0.002, 0.20	-0.89, 1.17
14. Spontaneous corrections	24	1,446	0.15	0.4325	0.2610	0.3449	0.19	0.4488	36	0.05, 0.25	-0.8, 0.76
15. Admitting lack of memory	28	1,883	0.28	0.3674	0.2519	0.2674	0.37	0.3480	47	0.18, 0.38	-0.07, 0.81
16. Doubts one's testimony	21	1,359	0.20	0.4082	0.2533	0.3201	0.26	0.4166	38	0.10, 0.30	-0.27, 0.80
17. Self-deprecation	7	526	-0.01	0.4670	0.2316	0.4055	-0.02	0.5276	25	-0.19, 0.17	-0.69, 0.66
18. Pardoning the perpetrator	3	284	0.007	0.2305	0.2062	0.1030	0.009	0.1341	80	-0.23, 0.24	-0.16, 0.18
19. Details characteristic offence	3	298	0.27	0.2147	0.2082	0.0524	0.36	0.0683	94	0.03, 0.50	0.27, 0.44
Reporting Style	3	357	0.41	0.0412	0.1874	0.0781	0.48	0.0909	85	0.19, 0.62	0.36, 0.59
Display insecurities	3	297	0.67	0.3069	0.2111	0.5122	0.78	0.5965	14	0.43, 0.90	0.01, 1.54
Providing reasons lack memory	4	447	0.15	0.2877	0.1902	0.2158	0.18	0.2514	44	-0.03, 0.32	-0.14, 0.50
Average (all criteria. 22)	39	2,841	0.25	0.4497	0.2336	0.3934	0.32	0.4933	37	0.17, 0.33	-0.31, 0.95
Total CBCA Score	26	1,770	0.47	0.5191	0.2470	0.4566	0.49	0.4786	23	0.37, 0.56	-0.12, 1.10



## 10. APÉNDICE





## 10.1. RESUMEN

La credibilidad del testimonio constituye un requerimiento esencial en la estimación de la prueba para la toma de decisiones judiciales. Esta tarea requiere de la participación de un perito experto al que se le atribuyen una serie de conocimientos y habilidades, así como una dilatada experiencia en la evaluación de la realidad del testimonio y en la identificación de sintomatología clínica consecuencia de la victimización de un delito. Un estudio elaborado por Arce y Fariña (2015) concluyó que la evaluación de la credibilidad del testimonio bajo mandato judicial permitió dotar de valor de prueba suficiente a la declaración en el 93.3% de los casos, y que muchos de los casos sobreseídos lo fueron por la usencia de dicho informe pericial. Por tanto, la práctica de la prueba pericial psicológica es especialmente relevante cuando los delitos cometidos tienen lugar en la intimidad y no se dispone de otras evidencias más que el testimonio de la víctima y agresor (i.e., delitos contra la libertad e indemnidad sexuales, violencia de género).

Si bien han surgido diversas técnicas para la evaluación de la credibilidad del testimonio, la más estudiada y que goza de mayor bagaje ha sido el Criteria-Based Content Analysis o CBCA, que forma parte de un instrumento más amplio denominado Statement Validity Assessment (SVA). El CBCA constituye el componente principal del SVA y se aplica a la entrevista de hechos una vez transcrita, cuyo objetivo consiste en analizar la calidad del contenido de la declaración a partir de la presencia/ausencia de 19 criterios de realidad. La hipótesis subyacente al CBCA recibe el nombre de *Hipótesis Undeutsch*, la cual establece que las memorias de hechos auto-experimentados difieren en contenido y calidad de las memorias de hechos inventados o fabricados. En este sentido, se sostiene que cuantos más criterios de realidad contenga la declaración, es más probable que ésta provenga de una memoria de hechos vividos, aunque de su ausencia no podemos concluir que la declaración sea fabricada o inventada, pues otras hipótesis alternativas podrían explicar la baja calidad del testimonio (e.g., la descripción de un hecho auto-experimentado simple en su naturaleza o un hecho puntual carente de detalles).

La investigación empírica ha mostrado heterogeneidad de opiniones y resultados aparentemente contradictorios respecto de la capacidad discriminativa de la técnica y su aplicabilidad en el campo forense. En términos prácticos, estas dudas se perfilan en si el CBCA y, por tanto, la Hipótesis Undeutsch, cumplen los criterios Daubert de admisibilidad de evidencia científica en los tribunales así como los requerimientos legales y jurisprudenciales para que la prueba, consistente en la declaración de la víctima, sea suficiente. No obstante, en ocasiones, la evaluación de la credibilidad del testimonio resulta insuficiente como prueba de cargo puesto que, para la acreditación de su credibilidad, el testimonio debe ser corroborado por circunstancias externas o periféricas (verosimilitud), concretándose en el establecimiento de la relación entre el delito y las posibles secuelas psicológicas consecuencia de la victimización. Por tanto, la evaluación forense debe contemplar no sólo el análisis de la credibilidad del testimonio mediante técnicas fiables y válidas, sino también cuantificar el daño ocasionado por la victimización de un delito sufrido en la esfera privada.

Ante este estado de la cuestión, nos planteamos los siguientes objetivos:

1. Poner a prueba la *Hipótesis Undeutsch*, a través de los criterios de realidad del CBCA, tanto en muestras de menores como en adultos, y en contextos distintos al abuso sexual. Es decir, evaluar la capacidad de los criterios de realidad para discriminar entre memorias vividas y fabricadas a partir de la estimación del tamaño del efecto (en qué medida las memorias de hechos auto-experimentados contienen más criterios de realidad que las fabricadas).
2. Analizar si los criterios de realidad del CBCA y, por ende, la Hipótesis Undeutsch, cumplen los *Daubert Standards* de admisibilidad de evidencia científica en el contexto judicial, con el objetivo de validar su uso en la práctica forense: 1) ¿Puede probarse la hipótesis científica? 2) Ha sido probada la hipótesis? 3) ¿Se conoce la tasa de error de la técnica? 4) ¿Ha sido sometida la hipótesis o la técnica a un proceso de revisión por pares y publicación? 5) ¿Se basa la hipótesis y/o técnica en una teoría generalmente aceptada como válida por la comunidad científica?
3. Estudiar la posible inclusión y combinación de los criterios del CBCA con criterios de realidad adicionales, que mejoren la eficacia de la técnica.
4. Estimar la probabilidad de desarrollar un trastorno internalizante (i.e., depresión y ansiedad) y cuantificar dichas secuelas, consecuencia de la victimización de un delito de naturaleza sexual.

Finalmente, se discuten las implicaciones de los resultados alcanzados en la evaluación forense, en relación con la práctica de la prueba psicológica y el estudio de su suficiencia en el proceso penal.

## MÉTODO

Se utilizó el procedimiento meta-analítico elaborado por Hunter y Schmidt (2015), como método estadístico para la acumulación sistemática de conocimiento científico. Esta técnica permite alcanzar conclusiones más precisas y generalizables gracias a la potencia estadística de sus resultados ya que aumenta significativamente el tamaño muestral. Además, permite identificar la influencia de determinadas variables en los resultados alcanzados y detectar áreas de incertidumbre. En consecuencia, el meta-análisis fue el método más adecuado para la consecución de los objetivos anteriormente expuestos.

Las etapas para la realización de los meta-análisis de experimentos (primero, segundo y cuarto meta-análisis de la presente tesis doctoral) en los que se estimó la *d* de Cohen (1988) como tamaño del efecto fueron: búsqueda exhaustiva y multimétodo con el fin de identificar estudios empíricos que pusieran a prueba la validez de los criterios de realidad del CBCA, en muestras de menores y adultos. Para ello, se recurrió a bases de datos internacionales de referencia científica, utilizando como motor de búsqueda las siguientes palabras clave en diferentes idiomas: Criteria-Based Content Analysis, CBCA, criterios de realidad, Statement Validity Assessment, credibilidad, análisis de contenido, testimonio, abuso sexual infantil, declaración, indicadores verbales, señales de engaño (entre otras). Asimismo, se ha continuado la búsqueda en meta-buscadores como Google Scholar; en redes sociales académicas (i.e., Researchgate, Academia.edu); se ha contactado con investigadores

destacados en la Psicología del Testimonio para solicitar estudios no publicados, y se han cotejado las listas de referencias de los estudios seleccionados para identificar investigaciones previas (*ancestry approach*).

Una vez identificados los potenciales estudios, se seleccionaron los que cumplían los criterios de inclusión y exclusión. Tras la selección de los estudios válidos de acuerdo con los objetivos de los meta-análisis, se clasificaron en una rejilla de codificación, teniendo en cuenta las variables que la literatura y los codificadores externos habían señalado como posibles moderadoras.

Los tamaños del efecto fueron tomados directamente de los estudios primarios cuando eran informados, en caso contrario, se calcularon a partir de las medias y desviaciones estándar/errores estándar. Cuando los resultados estaban expresados a partir de estadísticos de contraste, éstos se transformaron a  $d$  de Cohen utilizando las fórmulas de conversión de Rosenthal (1994) y Cohen (1988).

Con el objetivo de analizar la robustez de los efectos encontrados, se calculó el sesgo de publicación o *File Drawer Analysis* (FDA; McNatt, 2000), que permite conocer el número posible de estudios no detectados con efecto nulo que serían necesarios para reducir los efectos obtenidos en el meta-análisis a un efecto trivial (.05), próximo al cero.

Adicionalmente, se realizaron cálculos que exceden la finalidad propia del procedimiento meta-analítico pero la información que proporcionan era de interés: las diferencias entre tamaños del efecto ( $q$  de Cohen) con el fin de conocer la superioridad en la clasificación de declaraciones reales en unas determinadas circunstancias frente a otras (e.g., según el paradigma experimental); y la utilidad práctica de los resultados en el ámbito forense se estimó a partir de los estadísticos U1, Binomial Effect Size Display (BESD) y Probabilidad de Superioridad (PS).

Previamente a la realización de los meta-análisis, se estudió la existencia de posibles valores outliers o extremos. En caso afirmativo, se procedía a su eliminación, siempre y cuando no supusiera eliminar más de un 10% de los tamaños del efecto de cada meta-análisis (Tukey, 1960). Una vez seleccionados los estudios primarios que formarían parte del meta-análisis, se ponderó cada tamaño del efecto por el tamaño de la muestra y se corrigió por la falta de fiabilidad del criterio (influencia de artefactos). Para el cómputo de la fiabilidad del criterio se utilizaron las fiabilidades inter-jueces (i.e., codificación de los criterios de realidad como presentes o ausentes) informadas por los estudios primarios. Se estimó una fiabilidad promedio para los criterios de realidad, y otra para la puntuación total del CBCA, ya que es distinta a la de los criterios individuales (Horowitz et al., 1997), utilizando para ello la fórmula de Spearman-Brown. Además, se computó un intervalo de confianza (significación del tamaño del efecto) y un intervalo de credibilidad (generalización de los resultados). Asimismo, el porcentaje de varianza explicada por errores artificiales (% Var) fue estimado con el fin de analizar la heterogeneidad entre estudios y de la necesidad de buscar variables moderadoras que afecten a los resultados.

Finalmente, se mostró para cada meta-análisis una tabla de resultados con la siguiente información: número de tamaños del efecto ( $k$ ), tamaño de la muestra ( $n$ ), desviación estándar observada de  $d$  corregida por todos los artefactos ( $SD_{pre}$ ); desviación estándar observada de  $d$

después de eliminar la varianza debida a todos los artefactos ( $SD_{res}$ ); tamaño del efecto corregido por falta de fiabilidad del criterio ( $\delta$ ); desviación estándar de  $\delta$  ( $SD_{\delta}$ ); varianza explicada por todos los errores artificiales (%Var); intervalo de confianza para  $d$  al 95% (95%  $CI_d$ ); intervalo de credibilidad para  $\delta$  al 80% (80%  $CV_{\delta}$ ) o al 90% (90%  $CI_{\delta}$ ).

El proceso de búsqueda, identificación y selección de estudios para el meta-análisis de correlaciones fue el mismo, con la diferencia de que el efecto estimado en este caso fue la correlación de Pearson ( $r$ ) y que los criterios de selección de estudios fueron los pertinentes para la consecución de los objetivos establecidos en el meta-análisis. Se corrigió el tamaño del efecto no solo por el error de muestreo y el error de medida en la variable dependiente (falta de fiabilidad del criterio), sino también por el error de medida en la variable independiente (falta de fiabilidad del predictor, esto es, la medida de la victimización).

La tabla de resultados del meta-análisis de correlaciones, informaba de los siguientes indicadores: número de tamaños del efecto independientes ( $k$ ); tamaño muestral del grupo experimental ( $N_E$ ); tamaño muestral del grupo control ( $N_C$ ); tamaño muestral total ( $N_T$ ); correlación observada (validez observada) ponderada por el tamaño de la muestra ( $r_w$ ); desviación estándar de la correlación observada ( $SD_r$ ); correlación verdadera (validez operativa) corregida por falta de fiabilidad del predictor y del criterio, ( $\rho$ ); desviación estándar de la correlación verdadera ( $SD_{\rho}$ ); porcentaje de varianza explicada por los errores artificiales (%VE); intervalo de confianza al 95% (95%  $CI_r$ ); intervalo de credibilidad al 90% (90%  $CI_{\rho}$ ).

## RESULTADOS

**Artículo 1.** *Undeutsch Hypothesis and criteria Based Content Analysis. A meta-analytic review.*

**Artículo 2:** *Criteria-Based Content Analysis (CBCA) reality criteria in adults: A meta-analytic review.*

**Artículo 4:** *Análisis de contenido de la declaración en adultos. Un estudio meta-analítico.*

1. El tamaño del efecto corregido por la fiabilidad del criterio, tomando la puntuación total del CBCA, fue positivo, grande ( $\delta = 0.79$ ), significativo (el intervalo de confianza no contiene el valor cero) y generalizable en muestras de menores, mientras que en el meta-análisis de adultos fue de magnitud moderada ( $\delta = 0.56$ ) y no generalizable a otras poblaciones. El efecto obtenido fue significativamente superior en la muestra de menores.
2. En la muestra de menores, los efectos estimados con criterios de realidad del CBCA han sido positivos, significativos (con capacidad para discriminar entre memorias de hechos auto-experimentados y fabricados) pero no generalizables. La magnitud de los efectos osciló entre  $\delta = 0.17$  para el criterio ‘admisión de falta de memoria’, y  $\delta = 1.40$  para el criterio ‘detalles característicos de la agresión’.
3. En la muestra de adultos, se obtuvieron efectos positivos para los criterios de realidad del CBCA, con excepción del criterio ‘perdón al autor del delito’ ( $\delta = -0.02$ ) y significativos (salvo los criterios ‘auto-desaprobación’ y ‘perdón al autor



del delito' que no fueron productivos) pero no generalizables. El criterio con el tamaño del efecto más elevado fue el de 'cantidad de detalles',  $\delta = 0.71$ .

4. En general, los criterios de realidad discriminan por igual entre memorias vividas y fabricadas en muestras de menores y adultos. Solamente los criterios 'engranaje contextual', 'detalles superfluos', 'estado mental subjetivo' y 'detalles característicos de la ofensa' fueron significativamente más eficaces en la clasificación de las memorias vividas de menores.
5. Los criterios de realidad adicionales 'estilo de la declaración' (*issue related style*) y 'dar muestras de inseguridad', mostraron un efecto positivo ( $\delta = 0.48$  y  $\delta = 0.78$ , respectivamente), significativo y generalizable, mientras que el criterio 'dar explicaciones de la falta de memoria' no resultó productivo. Los criterios 'clichés' y 'repeticiones' mostraron un efecto negativo, aunque sólo fue significativo para éste último ( $\delta = -0.54$ ), apuntando a que se trata de un indicador de engaño.
6. Tomando el paradigma de investigación como variable moderadora, el efecto obtenido con los estudios de campo fue positivo, significativo y generalizable, en la población de menores (tamaño más que grande,  $\delta = 2.71$ ,  $p < .01$ ) y de adultos (magnitud moderada,  $\delta = 0.69$ ). Por su parte, los meta-análisis realizados con los estudios experimentales revelaron un efecto positivo y significativo para ambas muestras (menores,  $\delta = 0.56$ ; y adultos,  $\delta = 0.32$ ). No obstante, el efecto hallado en la población adulta no fue generalizable.
7. El estudio de la variable contexto (se tomaron los criterios significativos de los estudios de campo con casos reales de violencia sexual y de género) reveló un tamaño del efecto elevado y positivo, ( $\delta = 0.96$ ), significativo y generalizable. La diferencia entre tamaños del efecto mostró que la capacidad discriminativa de los criterios de realidad del CBCA es significativamente superior bajo estas circunstancias que tomando todos los estudios de campo,  $q_c = .168$ ,  $p < .05$  (0.96 vs. 0.69, respectivamente).

**Artículo 3.** *Psychological injury in victims of child sexual abuse. A meta-analytic review.*

8. La victimización de abuso sexual en la infancia o adolescencia conlleva una probabilidad aproximada de .70 en el desarrollo de sintomatología internalizante. El daño psicológico o sufrimiento emocional de las víctimas se concretó en un 28% en depresión y 31% en ansiedad.
9. Las mujeres víctimas tienen un 9% más de probabilidades de ser diagnosticadas de un trastorno de depresión que los hombres, cuantificándose el daño en un 41% frente al 10% en los hombres. La ratio de prevalencia de depresión y ansiedad fue aproximadamente 2 veces mayor, en mujeres y hombres víctimas frente a no víctimas.
10. La cuantificación de daño (trastorno de depresión) fue significativamente mayor para distimia (46%) que para depresión mayor (31%). Además, la prevalencia estimada de distimia y depresión mayor en víctimas de ASI/ASA, fue de 6 y 3 veces mayor, respectivamente, en comparación con las no víctimas.

11. Las víctimas de ASI/ASA informaron de secuelas de un trastorno de ansiedad, en torno al 43%.
12. El abuso sexual que incluye penetración, provoca un daño (depresión y ansiedad) significativamente mayor que los abusos sexuales sin contacto.

## DISCUSIÓN

**Artículo 1.** *Undeutsch Hypothesis and criteria Based Content Analysis. A meta-analytic review.*

1. Se valida la Hipótesis Undeutsch en memorias de menores de todas las edades y contextos (i.e., ajenos al abuso sexual) y, por extensión, los criterios de realidad del CBCA. Es decir, todos los criterios de realidad del CBCA discriminan entre memorias de hechos auto-experimentados y fabricados. Es más, las declaraciones basadas en hechos vividos contienen más criterios de realidad que las fabricadas.
2. El sistema categorial CBCA, aunque eficaz en los estudios de campo y de laboratorio, posee una capacidad discriminativa significativamente superior en los primeros. En consecuencia, los resultados obtenidos en los experimentales no son directamente generalizables a la práctica salvo que estén apoyados por estudios de campo.
3. Los criterios de realidad del componente cognitivo (criterios 1 al 13 del CBCA) mostraron unos tamaños del efecto más elevados en comparación con los motivacionales (criterios 14 al 18). Además, éstos no resultaron generalizables.
4. El criterio de realidad ‘detalles característicos de la agresión’ alcanzó el efecto más elevado. La explicación radicaría en la dificultad para inventar las características del propio evento.
5. En términos forenses, la puntuación total del CBCA clasifica correctamente (verdaderos positivos) al 68.5% de las memorias de hechos vividos, sin embargo, falla en su detección en el 31.5% de los casos (falsos negativos). Asimismo, la probabilidad de cometer falsos positivos, esto es, clasificar declaraciones falsas como verdaderas, es de .288, cuando en la práctica forense dicha probabilidad ha de ser 0, ya que transgrede el principio de presunción de inocencia (ninguna persona inocente debe estar en prisión, mientras que algunas personas que son culpables pueden estar fuera de prisión).
6. Teniendo en cuenta que el paradigma de investigación mostró ser una variable moderadora, se procedió al cálculo de la tasa de error con los estudios de experimentales y los de campo. El porcentaje de correcta clasificación tomando la puntuación total del CBCA fue del 90.2% en los estudios de campo frente al 63.5% en los estudios experimentales. Por tanto, los resultados de los estudios experimentales no pueden generalizarse directamente a la práctica, pues solamente gozan de validez aparente (Fariña et al., 1994; Konecni y Ebbesen, 1992). Asimismo, tomando los estudios de campo, la probabilidad de falsos positivos pasa a ser .029, aunque se desconoce el criterio de decisión para

alcanzar dicho resultado. En definitiva, no constituye una prueba válida puesto que la decisión estriba en una impresión global o juicio clínico (Arce & Fariña, 2013; Köhnken, 2004).

**Artículo 2:** *Criteria-Based Content Analysis (CBCA) reality criteria in adults: A meta-analytic review.*

**Artículo 4:** *Análisis de contenido de la declaración en adultos. Un estudio meta-analítico.*

7. Se validó la Hipótesis Undeutsch en adultos, y tanto en estudios experimentales (alta validez interna) como de campo (alta validez externa). No obstante, los criterios de realidad del CBCA clasificaron significativamente mejor las memorias vividas en estudios de campo que en los de diseños de simulación.
8. No todos los criterios de realidad fueron válidos (i.e., ‘auto-desaprobación’, ‘perdón al autor del delito’), ni generalizables (i.e., todos los criterios salvo ‘detalles incomprendidos relatados con precisión’ y ‘detalles característicos de la agresión’ que estaban afectados por un error de muestreo de segundo orden por lo que sus resultados no fueron válidos para esta estimación). Además, el criterio ‘perdón al autor del delito’ fue contrario a la Hipótesis Undeutsch (el tamaño del efecto medio fue negativo).
9. La Hipótesis Undeutsch fue validada no solamente para las memorias de víctimas, sino también para testigos y agresores. Sin embargo, los resultados extraídos del meta-análisis con muestras de agresores no pueden ser extrapolados a la práctica ya que los datos proceden de estudios exclusivamente experimentales. Además, en la evaluación forense de las memorias reales de agresores no puede establecerse la *ground truth*, ya que los agresores pueden, por ley, mentir, no declarar o negar los hechos que se les imputan, de forma alternativa a decir la verdad.
10. Los criterios de realidad del CBCA pueden complementarse con criterios adicionales de realidad (i.e., ‘estilo de la declaración’, ‘dar muestras de inseguridad’, y ‘dar explicaciones de la falta de memoria’). De los criterios adicionales estudiados, “repeticiones” y “clichés” mostraron una relación negativa con las memorias vividas, señalando que no se trata de criterios de realidad sino características o atributos de memoria no directamente relacionados con la realidad del evento.
11. Bajo la variable contexto, los criterios de realidad clasifican correctamente al 71.5% de las declaraciones reales de agresiones sexuales y violencia de género, mientras comete un 28.5% de falsos positivos.
12. Comparativamente, la eficacia discriminativa de la puntuación total del CBCA entre memorias vividas y fabricadas, fue significativamente superior en la muestra de menores que de adultos. En términos de criterios de realidad, los criterios ‘engranaje contextual’, ‘detalles superfluos’, ‘estado mental subjetivo’ y ‘detalles característicos de la agresión’ mejoran los ratios de correcta clasificación de las declaraciones reales en muestras de menores.

**Artículo 3.** *Psychological injury in victims of child sexual abuse. A meta-analytic review.*

13. Las víctimas de abuso sexual en la infancia (ASI) o en la adolescencia (ASA) tienen una probabilidad de .70 de padecer secuelas internalizantes, concretamente, .66 de depresión y .68 de ansiedad. La cuantificación del daño causado fue aproximadamente de un 30% en la secuela en general, depresión y ansiedad.
14. Concretamente, la probabilidad de desarrollar un daño crónico (distimia) es significativamente superior a la probabilidad de un daño más grave o severo (trastorno de depresión mayor).
15. Las víctimas de ASI/ASA tienen una probabilidad del 43% de desarrollar un trastorno de ansiedad (trastorno de ansiedad generalizada, fobia específica, fobia social y trastorno de pánico).
16. Las mujeres víctimas de ASI/ASA tienen un 9% más de probabilidades que los hombres de desarrollar depresión, así como un daño significativamente superior (41% en mujeres y 10% en hombres). Este hallazgo es consistente con estudios epidemiológicos y de prevalencia (APA, 2013; WHO, 2000) en los que se constata la vulnerabilidad de las mujeres en el desarrollo de secuelas internalizantes.
17. La victimización de ASI/ASA favorece la aparición de problemas de ajuste psicológico en la adultez (Chiesa, Larsen-Paya, Martino, & Trinchieri, 2016).

**IMPLICACIONES DE LOS RESULTADOS PARA LA PRÁCTICA FORENSE**

La admisibilidad de las pruebas científicas en el contexto judicial descansa en el cumplimiento de los Daubert Standards y criterios jurisprudenciales. Los meta-análisis realizados en esta tesis doctoral permiten dar respuesta a dichos criterios a través de la puesta a prueba de los criterios de realidad del CBCA:

1. ¿Puede probarse la hipótesis científica? ¿Ha sido probada la hipótesis? Numerosos estudios han sometido a prueba la Hipótesis Undeutsch, que subyace al CBCA y, los meta-análisis elaborados, han podido validarla y, por extensión, a los criterios de realidad del CBCA. Esto permite responder al primer y segundo criterio Daubert.
2. ¿Se conoce la tasa de error? Nuestros resultados han mostrado que la tasa de error en la clasificación de memorias reales (falsos positivos), fue de un 9.8% en casos reales de abuso sexual infantil y de 28.5% en casos reales de violencia de género y delitos de naturaleza sexual en muestras de adultos.
3. ¿Ha sido sometida la hipótesis/técnica a un proceso de revisión por pares y publicación? Muchos de los estudios que contrastan la técnica y la hipótesis han sido publicados y muchos han sido sometidos a un proceso de revisión por pares.

4. ¿Se basa la hipótesis y/o técnica en una teoría generalmente aceptada como válida por la comunidad científica? Aunque la comunidad científica no ha sido consultada como tal, la evidencia científica que procede de investigaciones primarias y, por consiguiente, los resultados meta-analíticos, proporcionan soporte a la hipótesis y a la técnica.

Aunque la evidencia científica valida la técnica, esta no es válida para la práctica forense ya que no proporciona un criterio de decisión estricto para la clasificación de las declaraciones, dejando que la decisión forense recaiga en una impresión del perito o en un juicio clínico. En definitiva, la decisión es semi-objetiva: basada en evidencia científica (objetiva) pero cuando se aplica a casos específicos ( $N = 1$ ) la decisión se sustenta en la impresión del experto (subjctiva), mientras que los tribunales requieren un criterio de decisión objetivo. Además, aunque la tasa de error de la técnica es conocida, no cumple el criterio jurisprudencial de presunción de inocencia puesto que en la evaluación forense, la técnica falla en la clasificación de falsos positivos entre un 9.8% en menores y un 28.5% en adultos. Asimismo, la técnica como sistema categorial de análisis de contenido no es válido, ya que sus categorías no son mutuamente excluyentes ni exhaustivas, tampoco objetivas y se desconoce si son homogéneas. Aunque generalmente los criterios de realidad fueron productivos (con capacidad discriminativa entre memorias vividas y fabricadas) algunas categorías no fueron pertinentes.

En resumen, el principal objetivo de esta tesis doctoral fue del proporcionar respuestas sobre la admisibilidad en los tribunales de sistemas categoriales para el análisis de la credibilidad. La técnica meta-analítica, con su labor integrativa de resultados empíricos, permitió alcanzar conclusiones generales. Los resultados apoyan la creación y validación de un sistema categorial metódico para la evaluación de la credibilidad, que debe estar compuesto por categorías de análisis de contenido válidas y fiables, y conformado por criterios de realidad específicos para cada contexto (i.e., delitos contra la libertad e indemnidad sexuales, violencia de género). De hecho, el CBCA no cumple algunos de los criterios legales (e.g., posee una elevada tasa de error de clasificación, falta de un criterio de decisión estricto) y requerimientos metodológicos (e.g., algunos criterios de realidad no son mutuamente excluyentes, y el sistema no es objetivo ni pertinente) en su forma actual. Asimismo, el método de evaluación debe poder estudiar la victimización a partir de la medida de daño psicológico como consecuencia de un delito sexual, con el objetivo de reparar el daño causado proporcionalmente al sufrimiento padecido, incluso de forma retrospectiva (el daño puede aparecer después de un largo período de tiempo e incluso cronificarse). Los resultados de los meta-análisis proporcionan a los jurados y jueces una estimación del alcance del daño, arrojando luz sobre el elevado coste psicológico de la victimización de delitos de naturaleza sexual. Por otro lado, además del diagnóstico de un trastorno internalizante, debe aplicarse la Escala de Evaluación de la Actividad Global (EEAG) al caso concreto con el objetivo de cuantificar el nivel general de sufrimiento de la persona evaluada (ajuste psicológico en las diferentes facetas de su vida diaria). Esta medida permite reparar el daño proporcionalmente al daño sufrido, y sería útil para ayudar a jueces en el alcance de sentencias judiciales más exactas y justas.



## 10.2. ARTÍCULOS







## ÍNDICE DE PUBLICACIONES DERIVADO DE LA TESIS

Amado, B. G., Arce, R., & Fariña, F. (2015). Undeutsch Hypothesis and criteria Based Content Analysis. A meta-analytic review. *The European Journal of Psychology Applied to Legal Context*, 7, 3-12. <https://doi.org/10.1016/j.ejpal.2014.11.002>

JCR Journal. Impact Factor: 1.000 (Q2 in JCR 2015)

Cite Score (2015): 1.35

Scimago Journal Rank: Q1 (Law), Q3 (Applied Psychology)

24 cites in Google, of which 8 in WOS

Amado, B. G., Arce, R., Fariña, F., & Vilariño, M. (2016). Criteria-Based Content Analysis (CBCA) reality criteria in adults: A meta-analytic review. *International Journal of Clinical and Health Psychology*, 16, 201-210. <https://doi.org/10.1016/j.ijchp.2016.01.002>

JCR Journal. Impact Factor: 2.217 (Q2 in JCR 2015)

Cite Score (2015): 2.28

Scimago Journal Rank: Q2 (Clinical Psychology)

7 cites in Google, of which 1 in WOS

Amado, B. G., Arce, R., & Herraiz, A. (2015). Psychological injury in victims of child sexual abuse. A meta-analytic review. *Psychosocial Intervention*, 24, 49-62. <https://doi.org/10.1016/j.psi.2015.03.002>

Journal indexed in Web of Science database, SCOPUS and with a FECYT Seal of Quality.

Cite Score (2015): 0.89

Scimago Journal Rank: Q3 (Applied Psychology, Social Psychology, Developmental and Educational Psychology)

9 cites in Google

Amado, B. G., Arce, R., & Vilariño, M. (2016). Análisis de contenido de la declaración en adultos. Un estudio meta-analítico. En R. Arce, B. G. Amado, y M. Vilariño (Eds.). *Libro de Actas del IX Congreso Internacional de Psicología Jurídica y Forense*. Santiago de Compostela. ISBN: 978-84-8408-907-0

[https://www.researchgate.net/profile/Manuel\\_Vilarino/publication/296485509\\_Content\\_analysis\\_of\\_offenders\\_statements\\_A\\_meta-analytic\\_review\\_Analisis\\_de\\_contenido\\_en\\_declaraciones\\_de\\_agresores\\_Una\\_revision\\_meta-analitica/links/56d5de2308aeb4638ac684d.pdf](https://www.researchgate.net/profile/Manuel_Vilarino/publication/296485509_Content_analysis_of_offenders_statements_A_meta-analytic_review_Analisis_de_contenido_en_declaraciones_de_agresores_Una_revision_meta-analitica/links/56d5de2308aeb4638ac684d.pdf)